# Mathematics for Computer Science

revised Wednesday 4th January, 2012, 13:53

Eric Lehman
Google Inc.

F Thomson Leighton
Department of Mathematics
and the Computer Science and AI Laboratory,
Massachussetts Institute of Technology;
Akamai Technologies

Albert R Meyer
Department of Electrical Engineering and Computer Science
and the Computer Science and AI Laboratory,
Massachussetts Institute of Technology

# III Counting

# Introduction

Counting is useful in computer science for several reasons:

- Determining the time and storage required to solve a computational problem —a central objective in computer science —often comes down to solving a counting problem.
- Counting is the basis of probability theory, which plays a central role in all sciences, including computer science.
- Two remarkable proof techniques, the "pigeonhole principle" and "combinatorial proof," rely on counting.

Counting seems easy enough: 1, 2, 3, 4, etc. This direct approach works well for counting simple things —like your toes —and may be the only approach for extremely complicated things with no identifiable structure. However, subtler methods can help you count many things in the vast middle ground, such as:

- The number of different ways to select a dozen doughnuts when there are five varieties available.
- The number of 16-bit numbers with exactly 4 ones.

Perhaps surprisingly, but certainly not coincidentally, these two numbersa are the same: 1820.

We begin our study of counting in Chapter 14 with a collection of rules and methods for finding closed-form expressions for commonly-occurring sums and products such as $\sum_{i=1}^{n} x^i$ and $n! = \prod_{i=1}^{n} i$. We also introduce asymptotic notations such as $\sim$, $O$, and $\Theta$ that are commonly used in computer science to express

the how a quantity such as the running time of a program grows with the size of the input.

Chapter 15 describes the most basic rules for determining the cardinality of a set. These rules are actually theorems, but our focus won't be on their proofs *per se* —our objective is to teach you simple counting as a practical skill, like integration.

But counting can be tricky, and people make counting mistakes all the time, so a crucial part of counting skill is being able to verify a counting argument. Sometimes this can be done simply by finding an alternative way to count and then comparing answers —they better agree. But most elementary counting arguments reduce to finding a bijection between objects to be counted and easy-to-count sequences. The chapter shows how explicitly defining these bijections —and verifying that they are bijections —is another useful way to verify counting arguments. The material in Chapter 15 is simple yet powerful, and it provides a great tool set for use in your future career.

# 14 Sums and Asymptotics

Sums and products arise regularly in the analysis of algorithms, financial applications, physical problems, and probabilistic systems. For example, according to Theorem 2.2.1,

$$1 + 2 + 3 + \cdots + n = \frac{n(n+1)}{2}. \tag{14.1}$$

Of course the lefthand sum could be expressed concisely as a subscripted summation

$$\sum_{i=1}^{n} i,$$

but the right hand expression $n(n+1)/2$ is not only concise, it is also easier to evaluate, and it more clearly reveals properties such as the growth rate of the sum. Expressions like $n(n+1)/2$ that do not make use of subscripted summations or products —or those handy but sometimes troublesome dots —are called *closed forms*.

Another example is the closed form for a *geometric sum*

$$1 + x + x^2 + x^3 + \cdots + x^n = \frac{1 - x^{n+1}}{1 - x} \tag{14.2}$$

given in Problem 6.2. The sum as described on the left hand side of (14.2) involves $n$ additions and $1 + 2 + \cdots + (n-1) = (n-1)n/2$ multiplications, but its closed form on the right hand side can be evaluated using fast exponentiation with at most $2 \log n$ multiplications and a couple of substractions. Also, the closed form makes the growth and limiting behavior of the sum much more apparent.

Equations (14.1) and (14.2) were easy to verify by induction, but, as is often the case, the proofs by induction gave no hint about how these formulas were found in the first place. Finding them is part math and part art, which we'll start examining in this chapter.

A first motivating example will be figuring out the value of the annuity. The value will be a large and nasty-looking sum. We will then describe several methods for finding closed forms for several sorts of sums, including the annuity sums. In some cases, a closed form for a sum may not exist and so we will provide a general method for finding closed forms for good upper and lower bounds on the sum.

The methods we develop for sums will also work for products since any product can be converted into a sum by taking a logarithm of the product. As an example,

we will use this approach to find a good closed-form approximation to the *factorial function*

$$n! ::= 1 \cdot 2 \cdot 3 \cdots n.$$

We conclude the chapter with a discussion of asymptotic notation. Asymptotic notation is often used to bound the error terms when there is no exact closed form expression for a sum or product. It also provides a convenient way to express the growth rate or order of magnitude of a sum or product.

## 14.1 The Value of an Annuity

Would you prefer a million dollars today or $50,000 a year for the rest of your life? On the one hand, instant gratification is nice. On the other hand, the *total dollars* received at $50K per year is much larger if you live long enough.

Formally, this is a question about the value of an annuity. An *annuity* is a financial instrument that pays out a fixed amount of money at the beginning of every year for some specified number of years. In particular, an $n$-year, $m$-payment annuity pays $m$ dollars at the start of each year for $n$ years. In some cases, $n$ is finite, but not always. Examples include lottery payouts, student loans, and home mortgages. There are even Wall Street people who specialize in trading annuities.[1]

A key question is, "What is an annuity worth?" For example, lotteries often pay out jackpots over many years. Intuitively, $50,000 a year for 20 years ought to be worth less than a million dollars right now. If you had all the cash right away, you could invest it and begin collecting interest. But what if the choice were between $50,000 a year for 20 years and a *half* million dollars today? Now it is not clear which option is better.

### 14.1.1 The Future Value of Money

In order to answer such questions, we need to know what a dollar paid out in the future is worth today. To model this, let's assume that money can be invested at a fixed annual interest rate $p$. We'll assume an 8% rate[2] for the rest of the discussion.

Here is why the interest rate $p$ matters. Ten dollars invested today at interest rate $p$ will become $(1+p)\cdot 10 = 10.80$ dollars in a year, $(1+p)^2 \cdot 10 \approx 11.66$ dollars

[1] Such trading ultimately led to the subprime mortgage disaster in 2008–2009. We'll talk more about that in a later chapter.

[2] U.S. interest rates have dropped steadily for several years, and ordinary bank deposits now earn around 1.0%. But just a few years ago the rate was 8%; this rate makes some of our examples a little more dramatic. The rate has been as high as 17% in the past thirty years.

in two years, and so forth. Looked at another way, ten dollars paid out a year from now is only really worth $1/(1+p)\cdot 10 \approx 9.26$ dollars today. The reason is that if we had the \$9.26 today, we could invest it and would have \$10.00 in a year anyway. Therefore, $p$ determines the value of money paid out in the future.

So for an $n$-year, $m$-payment annuity, the first payment of $m$ dollars is truly worth $m$ dollars. But the second payment a year later is worth only $m/(1+p)$ dollars. Similarly, the third payment is worth $m/(1+p)^2$, and the $n$-th payment is worth only $m/(1+p)^{n-1}$. The total value, $V$, of the annuity is equal to the sum of the payment values. This gives:

$$\begin{aligned} V &= \sum_{i=1}^{n} \frac{m}{(1+p)^{i-1}} \\ &= m \cdot \sum_{j=0}^{n-1} \left(\frac{1}{1+p}\right)^{j} && \text{(substitute } j = i-1\text{)} \\ &= m \cdot \sum_{j=0}^{n-1} x^{j} && \text{(substitute } x = 1/(1+p)\text{).} \end{aligned} \tag{14.3}$$

The goal of the preceding substitutions was to get the summation into the form of a simple geometric sum. This leads us to an explanation of a way you could have discovered the closed form (14.2) in the first place using the *Perturbation Method*.

### 14.1.2 The Perturbation Method

Given a sum that has a nice structure, it is often useful to "perturb" the sum so that we can somehow combine the sum with the perturbation to get something much simpler. For example, suppose

$$S = 1 + x + x^2 + \cdots + x^n.$$

An example of a perturbation would be

$$xS = x + x^2 + \cdots + x^{n+1}.$$

The difference between $S$ and $xS$ is not so great, and so if we were to subtract $xS$ from $S$, there would be massive cancellation:

$$\begin{aligned} S &= 1 + x + x^2 + x^3 + \cdots + x^n \\ -xS &= \quad - x - x^2 - x^3 - \cdots - x^n - x^{n+1}. \end{aligned}$$

The result of the subtraction is

$$S - xS = 1 - x^{n+1}.$$

Solving for $S$ gives the desired closed-form expression in equation 14.2, namely,

$$S = \frac{1 - x^{n+1}}{1 - x}.$$

We'll see more examples of this method when we introduce *generating functions* in a later chapter.

### 14.1.3 A Closed Form for the Annuity Value

Using equation 14.2, we can derive a simple formula for $V$, the value of an annuity that pays $m$ dollars at the start of each year for $n$ years.

$$V = m\left(\frac{1 - x^n}{1 - x}\right) \qquad \text{(by equations 14.3 and 14.2)} \qquad (14.4)$$

$$= m\left(\frac{1 + p - (1/(1+p))^{n-1}}{p}\right) \qquad \text{(substituting } x = 1/(1+p)\text{)}. \qquad (14.5)$$

Equation 14.5 is much easier to use than a summation with dozens of terms. For example, what is the real value of a winning lottery ticket that pays \$50,000 per year for 20 years? Plugging in $m = \$50{,}000$, $n = 20$, and $p = 0.08$ gives $V \approx \$530{,}180$. So because payments are deferred, the million dollar lottery is really only worth about a half million dollars! This is a good trick for the lottery advertisers.

### 14.1.4 Infinite Geometric Series

The question we began with was whether you would prefer a million dollars today or \$50,000 a year for the rest of your life. Of course, this depends on how long you live, so optimistically assume that the second option is to receive \$50,000 a year *forever*. This sounds like infinite money! But we can compute the value of an annuity with an infinite number of payments by taking the limit of our geometric sum in equation 14.2 as $n$ tends to infinity.

**Theorem 14.1.1.** *If* $|x| < 1$, *then*

$$\sum_{i=0}^{\infty} x^i = \frac{1}{1 - x}.$$

*Proof.*

$$\begin{aligned}\sum_{i=0}^{\infty} x^i &::= \lim_{n\to\infty} \sum_{i=0}^{n} x^i \\ &= \lim_{n\to\infty} \frac{1-x^{n+1}}{1-x} && \text{(by equation 14.2)} \\ &= \frac{1}{1-x}.\end{aligned}$$

The final line follows from that fact that $\lim_{n\to\infty} x^{n+1} = 0$ when $|x| < 1$. ■

In our annuity problem, $x = 1/(1+p) < 1$, so Theorem 14.1.1 applies, and we get

$$\begin{aligned}V &= m \cdot \sum_{j=0}^{\infty} x^j && \text{(by equation 14.3)} \\ &= m \cdot \frac{1}{1-x} && \text{(by Theorem 14.1.1)} \\ &= m \cdot \frac{1+p}{p} && (x = 1/(1+p)).\end{aligned}$$

Plugging in $m = \$50{,}000$ and $p = 0.08$, we see that the value $V$ is only \$675,000. Amazingly, a million dollars today is worth much more than \$50,000 paid every year forever! Then again, if we had a million dollars today in the bank earning 8% interest, we could take out and spend \$80,000 a year forever. So on second thought, this answer really isn't so amazing.

### 14.1.5 Examples

Equation 14.2 and Theorem 14.1.1 are incredibly useful in computer science.

Here are some other common sums that can be put into closed form using equa-

tion 14.2 and Theorem 14.1.1:

$$1 + 1/2 + 1/4 + \cdots = \sum_{i=0}^{\infty} \left(\frac{1}{2}\right)^i = \frac{1}{1-(1/2)} = 2 \tag{14.6}$$

$$0.99999\cdots = 0.9\sum_{i=0}^{\infty} \left(\frac{1}{10}\right)^i = 0.9\left(\frac{1}{1-1/10}\right) = 0.9\left(\frac{10}{9}\right) = 1 \tag{14.7}$$

$$1 - 1/2 + 1/4 - \cdots = \sum_{i=0}^{\infty} \left(\frac{-1}{2}\right)^i = \frac{1}{1-(-1/2)} = \frac{2}{3} \tag{14.8}$$

$$1 + 2 + 4 + \cdots + 2^{n-1} = \sum_{i=0}^{n-1} 2^i = \frac{1-2^n}{1-2} = 2^n - 1 \tag{14.9}$$

$$1 + 3 + 9 + \cdots + 3^{n-1} = \sum_{i=0}^{n-1} 3^i = \frac{1-3^n}{1-3} = \frac{3^n-1}{2} \tag{14.10}$$

If the terms in a geometric sum grow smaller, as in equation 14.6, then the sum is said to be *geometrically decreasing*. If the terms in a geometric sum grow progressively larger, as in equations 14.9 and 14.10, then the sum is said to be *geometrically increasing*. In either case, the sum is usually approximately equal to the term in the sum with the greatest absolute value. For example, in equations 14.6 and 14.8, the largest term is equal to 1 and the sums are 2 and 2/3, both relatively close to 1. In equation 14.9, the sum is about twice the largest term. In equation 14.10, the largest term is $3^{n-1}$ and the sum is $(3^n - 1)/2$, which is only about a factor of 1.5 greater. You can see why this rule of thumb works by looking carefully at equation 14.2 and Theorem 14.1.1.

### 14.1.6 Variations of Geometric Sums

We now know all about geometric sums —if you have one, life is easy. But in practice one often encounters sums that cannot be transformed by simple variable substitutions to the form $\sum x^i$.

A non-obvious, but useful way to obtain new summation formulas from old ones is by differentiating or integrating with respect to $x$. As an example, consider the following sum:

$$\sum_{i=1}^{n-1} i x^i = x + 2x^2 + 3x^3 + \cdots + (n-1)x^{n-1}$$

This is not a geometric sum, since the ratio between successive terms is not fixed, and so our formula for the sum of a geometric sum cannot be directly applied. But

differentiating equation 14.2 leads to:

$$\frac{d}{dx}\left(\sum_{i=0}^{n-1} x^i\right) = \frac{d}{dx}\left(\frac{1-x^n}{1-x}\right). \tag{14.11}$$

The left-hand side of equation 14.11 is simply

$$\sum_{i=0}^{n-1} \frac{d}{dx}(x^i) = \sum_{i=0}^{n-1} i x^{i-1}.$$

The right-hand side of equation 14.11 is

$$\begin{aligned}\frac{-nx^{n-1}(1-x)-(-1)(1-x^n)}{(1-x)^2} &= \frac{-nx^{n-1}+nx^n+1-x^n}{(1-x)^2}\\ &= \frac{1-nx^{n-1}+(n-1)x^n}{(1-x)^2}.\end{aligned}$$

Hence, equation 14.11 means that

$$\sum_{i=0}^{n-1} i x^{i-1} = \frac{1-nx^{n-1}+(n-1)x^n}{(1-x)^2}.$$

Incidentally, Problem 14.2 shows how the perturbation method could also be applied to derive this formula.

Often, differentiating or integrating messes up the exponent of $x$ in every term. In this case, we now have a formula for a sum of the form $\sum i x^{i-1}$, but we want a formula for the series $\sum i x^i$. The solution is simple: multiply by $x$. This gives:

$$\sum_{i=1}^{n-1} i x^i = \frac{x-nx^n+(n-1)x^{n+1}}{(1-x)^2} \tag{14.12}$$

and we have the desired closed-form expression for our sum[3]. It's a little complicated looking, but it's easier to work with than the sum.

Notice that if $|x| < 1$, then this series converges to a finite value even if there are infinitely many terms. Taking the limit of equation 14.12 as $n$ tends infinity gives the following theorem:

[3]Since we could easily have made a mistake in the calculation, it is always a good idea to go back and validate a formula obtained this way with a proof by induction.

**Theorem 14.1.2.** *If* $|x| < 1$, *then*

$$\sum_{i=1}^{\infty} i x^i = \frac{x}{(1-x)^2}. \tag{14.13}$$

As a consequence, suppose that there is an annuity that pays $im$ dollars at the *end* of each year $i$ forever. For example, if $m = \$50{,}000$, then the payouts are \$50,000 and then \$100,000 and then \$150,000 and so on. It is hard to believe that the value of this annuity is finite! But we can use Theorem 14.1.2 to compute the value:

$$\begin{aligned} V &= \sum_{i=1}^{\infty} \frac{im}{(1+p)^i} \\ &= m \cdot \frac{1/(1+p)}{(1-\frac{1}{1+p})^2} \\ &= m \cdot \frac{1+p}{p^2}. \end{aligned}$$

The second line follows by an application of Theorem 14.1.2. The third line is obtained by multiplying the numerator and denominator by $(1+p)^2$.

For example, if $m = \$50{,}000$, and $p = 0.08$ as usual, then the value of the annuity is $V = \$8{,}437{,}500$. Even though the payments increase every year, the increase is only additive with time; by contrast, dollars paid out in the future decrease in value exponentially with time. The geometric decrease swamps out the additive increase. Payments in the distant future are almost worthless, so the value of the annuity is finite.

The important thing to remember is the trick of taking the derivative (or integral) of a summation formula. Of course, this technique requires one to compute nasty derivatives correctly, but this is at least theoretically possible!

## 14.2 Sums of Powers

In Chapter 6, we verified the formula (14.1), but the source of this formula is still a mystery. Sure, we can prove it is true using well ordering or induction, but where did the expression on the right come from in the first place? Even more inexplicable is the closed form expression for the sum of consecutive squares:

$$\sum_{i=1}^{n} i^2 = \frac{(2n+1)(n+1)n}{6}. \tag{14.14}$$

It turns out that there is a way to derive these expressions, but before we explain it, we thought it would be fun[4] to show you how Gauss is supposed to have proved equation 14.1 when he was a young boy.

Gauss's idea is related to the perturbation method we used in Section 14.1.2. Let

$$S = \sum_{i=1}^{n} i.$$

Then we can write the sum in two orders:

$$\begin{aligned} S &= 1 + \quad 2 \quad + \ldots + (n-1) + n, \\ S &= n + (n-1) + \ldots + \quad 2 \quad + 1. \end{aligned}$$

Adding these two equations gives

$$\begin{aligned} 2S &= (n+1) + (n+1) + \cdots + (n+1) + (n+1) \\ &= n(n+1). \end{aligned}$$

Hence,

$$S = \frac{n(n+1)}{2}.$$

Not bad for a young child —Gauss showed some potential....

Unfortunately, the same trick does not work for summing consecutive squares. However, we can observe that the result might be a third-degree polynomial in $n$, since the sum contains $n$ terms that average out to a value that grows quadratically in $n$. So we might guess that

$$\sum_{i=1}^{n} i^2 = an^3 + bn^2 + cn + d.$$

If the guess is correct, then we can determine the parameters $a$, $b$, $c$, and $d$ by plugging in a few values for $n$. Each such value gives a linear equation in $a$, $b$, $c$, and $d$. If we plug in enough values, we may get a linear system with a unique solution. Applying this method to our example gives:

$$\begin{aligned} n = 0 \quad &\text{implies} \quad 0 = d \\ n = 1 \quad &\text{implies} \quad 1 = a + b + c + d \\ n = 2 \quad &\text{implies} \quad 5 = 8a + 4b + 2c + d \\ n = 3 \quad &\text{implies} \quad 14 = 27a + 9b + 3c + d. \end{aligned}$$

[4] OK, our definition of "fun" may be different than yours.

Solving this system gives the solution $a = 1/3$, $b = 1/2$, $c = 1/6$, $d = 0$. Therefore, *if* our initial guess at the form of the solution was correct, then the summation is equal to $n^3/3 + n^2/2 + n/6$, which matches equation 14.14.

The point is that if the desired formula turns out to be a polynomial, then once you get an estimate of the *degree* of the polynomial, all the coefficients of the polynomial can be found automatically.

**Be careful!** This method lets you discover formulas, but it doesn't guarantee they are right! After obtaining a formula by this method, it's important to go back and *prove* it using induction or some other method, because if the initial guess at the solution was not of the right form, then the resulting formula will be completely wrong! A later chapter will describe a method based on generating functions that does not require any guessing at all.

## 14.3 Approximating Sums

Unfortunately, it is not always possible to find a closed-form expression for a sum. For example, consider the sum

$$S = \sum_{i=1}^{n} \sqrt{i}.$$

No closed form expression is known for $S$.

In such cases, we need to resort to approximations for $S$ if we want to have a closed form. The good news is that there is a general method to find closed-form upper and lower bounds that work for most any sum. Even better, the method is simple and easy to remember. It works by replacing the sum by an integral and then adding either the first or last term in the sum.

**Definition 14.3.1.** A function $f : \mathbb{R}^+ \to \mathbb{R}^+$ is *strictly increasing* when

$$x < y \text{ IMPLIES } f(x) < f(y),$$

and it is *weakly increasing*[5] when

$$x < y \text{ IMPLIES } f(x) \leq f(y).$$

[5]Weakly increasing functions are usually called *nondecreasing* functions. We will avoid this terminology to prevent confusion between being a nondecreasing function and the much weaker property of *not* being a decreasing function.

Similarly, $f$ is *strictly decreasing* when

$$x < y \text{ IMPLIES } f(x) > f(y),$$

and it is *weakly decreasing*[6] when

$$x < y \text{ IMPLIES } f(x) \geq f(y).$$

For example, $2^x$ and $\sqrt{x}$ are strictly increasing functions, while $\max x, 2$ and $\lceil x \rceil$ are weakly increasing functions. The functions $1/x$ and $2^{-x}$ are strictly decreasing, while $\min 1/x, 1/2$ and $\lfloor 1/x \rfloor$ are weakly decreasing.

**Theorem 14.3.2.** *Let $f : \mathbb{R}^+ \to \mathbb{R}^+$ be a weakly increasing function. Define*

$$S ::= \sum_{i=1}^{n} f(i) \tag{14.15}$$

*and*

$$I ::= \int_1^n f(x)\,dx.$$

*Then*

$$I + f(1) \leq S \leq I + f(n). \tag{14.16}$$

*Similarly, if $f$ is weakly decreasing, then*

$$I + f(n) \leq S \leq I + f(1).$$

*Proof.* Suppose $f : \mathbb{R}^+ \to \mathbb{R}^+$ is weakly increasing. The value of the sum $S$ in (14.15) is the sum of the areas of $n$ unit-width rectangles of heights $f(1), f(2), \ldots, f(n)$. This area of these rectangles is shown shaded in Figure 14.1.

The value of

$$I = \int_1^n f(x)\,dx$$

is the shaded area under the curve of $f(x)$ from 1 to $n$ shown in Figure 14.2.

Comparing the shaded regions in Figures 14.1 and 14.2 shows that $S$ is at least $I$ plus the area of the leftmost rectangle. Hence,

$$S \geq I + f(1) \tag{14.17}$$

This is the lower bound for $S$ given in (14.16).

To derive the upper bound for $S$ given in (14.16), we shift the curve of $f(x)$ from 1 to $n$ one unit to the left as shown in Figure 14.3.

[6] Weakly decreasing functions are usually called *nonincreasing*.

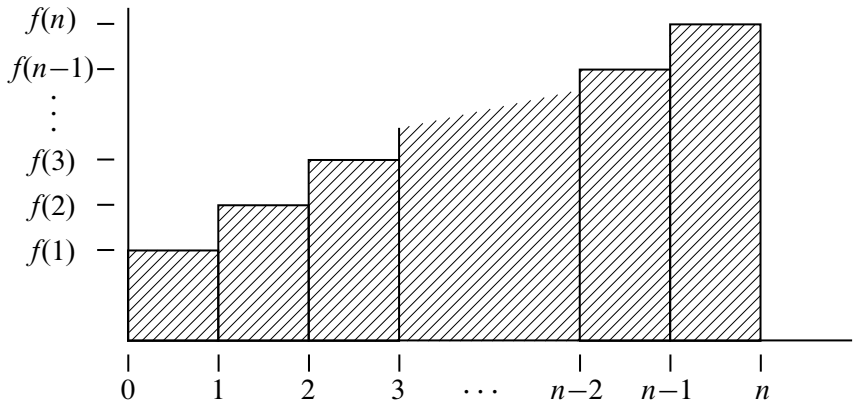



**Figure 14.1** The area of the $i$th rectangle is $f(i)$. The shaded region has area $\sum_{i=1}^{n} f(i)$.

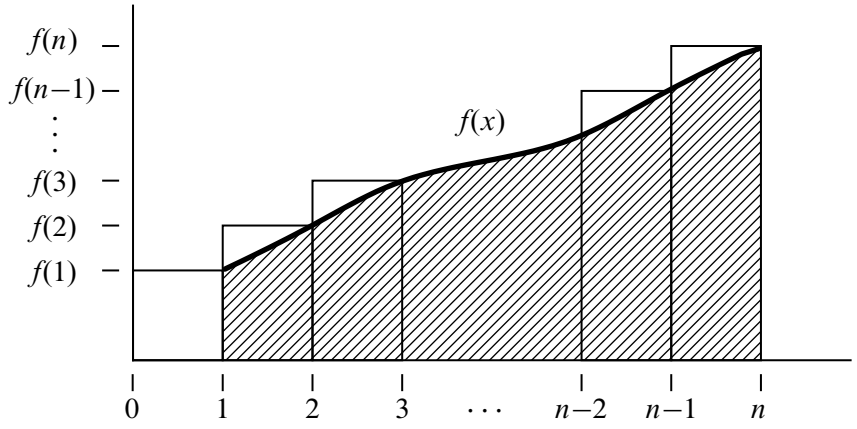



**Figure 14.2** The shaded area under the curve of $f(x)$ from 1 to $n$ (shown in bold) is $I = \int_1^n f(x)\,dx$.

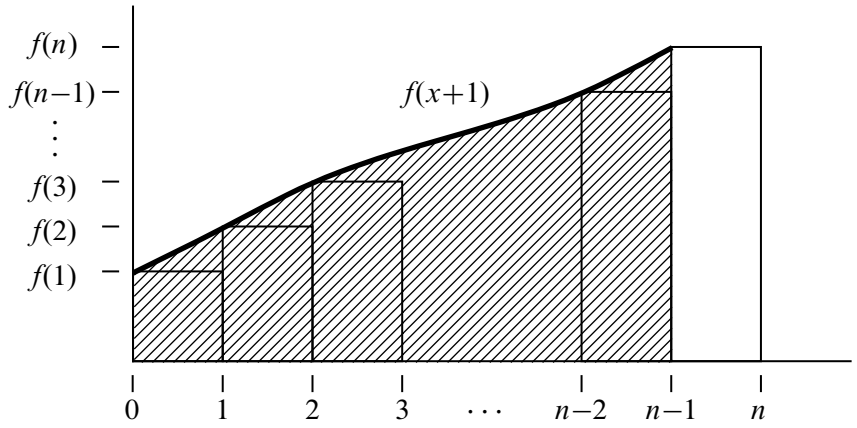



**Figure 14.3** This curve is the same as the curve in Figure 14.2 shifted left by 1.

Comparing the shaded regions in Figures 14.1 and 14.3 shows that $S$ is at most $I$ plus the area of the rightmost rectangle. That is,

$$S \leq I + f(n),$$

which is the upper bound for $S$ given in (14.16).

The very similar argument for the weakly decreasing case is left to Problem 14.7. ■

Theorem 14.3.2 provides good bounds for most sums. At worst, the bounds will be off by the largest term in the sum. For example, we can use Theorem 14.3.2 to bound the sum

$$S = \sum_{i=1}^{n} \sqrt{i}$$

as follows.

We begin by computing

$$\begin{aligned} I &= \int_1^n \sqrt{x}\, dx \\ &= \frac{x^{3/2}}{3/2}\Bigg|_1^n \\ &= \frac{2}{3}(n^{3/2} - 1). \end{aligned}$$

We then apply Theorem 14.3.2 to conclude that

$$\frac{2}{3}(n^{3/2}-1)+1 \;\le\; S \;\le\; \frac{2}{3}(n^{3/2}-1)+\sqrt{n}$$

and thus that

$$\frac{2}{3}n^{3/2}+\frac{1}{3} \;\le\; S \;\le\; \frac{2}{3}n^{3/2}+\sqrt{n}-\frac{2}{3}.$$

In other words, the sum is very close to $\frac{2}{3}n^{3/2}$.

We'll be using Theorem 14.3.2 extensively going forward. At the end of this chapter, we will also introduce some notation that expresses phrases like "the sum is very close to" in a more precise mathematical manner. But first, we'll see how Theorem 14.3.2 can be used to resolve a classic paradox in structural engineering.

## 14.4 Hanging Out Over the Edge

Suppose we have $n$ identical unit length rectangular blocks that are uniformly weighted. We want to stack them one on top of the next on a table as shown in Figure 14.4. Is there some value of $n$ for which it is possible to arrange the stack so that one of the blocks hangs out completely over the edge of the table without having the stack fall over? (You are not allowed to use glue or otherwise hold the stack in position.)

Most people's first response to this question —sometimes also their second and third responses —is "No. No block will ever get completely past the edge of the table." But in fact, if $n$ is large enough, you can get the top block to stick out as far as you want: one block-length, two block-lengths, any number of block-lengths!

### 14.4.1 Stability

A stack of blocks is said to be *stable* if it will not fall over of its own accord. For example, the stack illustrated in Figure 14.4 is not stable because the top block is sure to fall over. This is because the center or mass of the top block is hanging out over air.

In general, a stack of $n$ blocks will be stable if and only if the center of mass of the top $i$ blocks sits over the $(i+1)$st block for $i = 1, 2, \ldots, n-1$, and over the table for $i = n$.

We define the *overhang* of a stable stack to be the distance between the edge of the table and the rightmost end of the rightmost block in the stack. Our goal is thus to maximize the overhang of a stable stack.

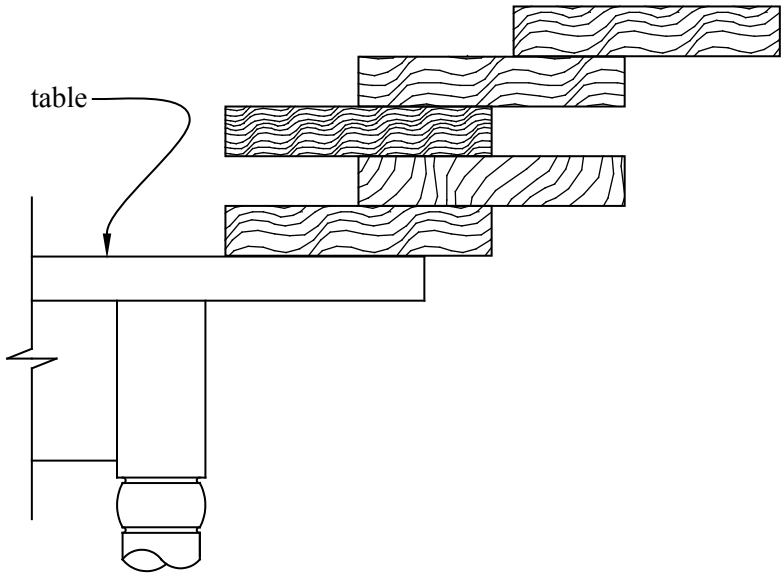



**Figure 14.4** A stack of 5 identical blocks on a table. The top block is hanging out over the edge of the table, but if you try stacking the blocks this way, the stack will fall over.

For example, the maximum possible overhang for a single block is $1/2$. That is because the center of mass of a single block is in the middle of the block (which is distance $1/2$ from the right edge of the block). If we were to place the block so that its right edge is more than $1/2$ from the edge of the table, the center of mass would be over air and the block would tip over. But we can place the block so the center of mass is at the edge of the table, thereby achieving overhang $1/2$. This position is illustrated in Figure 14.5.

In general, the overhang of a stack of blocks is maximized by sliding the entire stack rightward until its center of mass is at the edge of the table. The overhang will then be equal to the distance between the center of mass of the stack and the rightmost edge of the rightmost block. We call this distance the *spread* of the stack. Note that the spread does not depend on the location of the stack on the table—it is purely a property of the blocks in the stack. Of course, as we just observed, the maximum possible overhang is equal to the maximum possible spread. This relationship is illustrated in Figure 14.6.

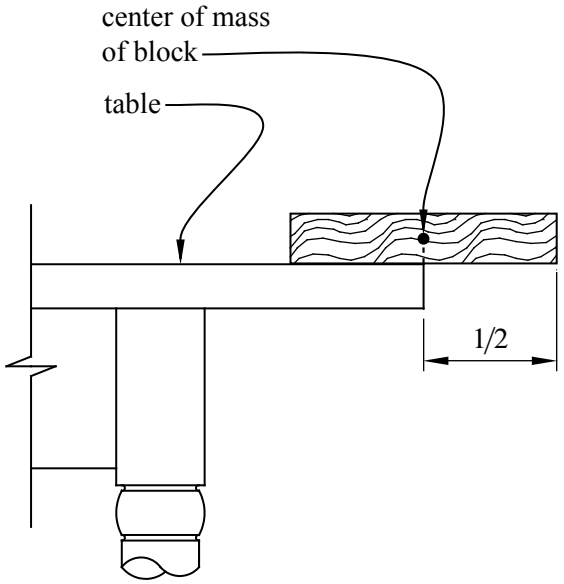



**Figure 14.5** One block can overhang half a block length.

### 14.4.2 A Recursive Solution

Our goal is to find a formula for the maximum possible spread $S_n$ that is achievable with a stable stack of $n$ blocks.

We already know that $S_1 = 1/2$ since the right edge of a single block with length 1 is always distance $1/2$ from its center of mass. Let's see if we can use a recursive approach to determine $S_n$ for all $n$. This means that we need to find a formula for $S_n$ in terms of $S_i$ where $i < n$.

Suppose we have a stable stack $\mathcal{S}$ of $n$ blocks with maximum possible spread $S_n$. There are two cases to consider depending on where the rightmost block is in the stack.

**Case 1:** *The rightmost block in $\mathcal{S}$ is the bottom block.* Since the center of mass of the top $n-1$ blocks must be over the bottom block for stability, the spread is maximized by having the center of mass of the top $n-1$ blocks be directly over the *left* edge of the bottom block. In this case the center of mass of $\mathcal{S}$ is[7]

$$\frac{(n-1)\cdot 1 + (1)\cdot \frac{1}{2}}{n} = 1 - \frac{1}{2n}$$

[7]The center of mass of a stack of blocks is the average of the centers of mass of the individual blocks.

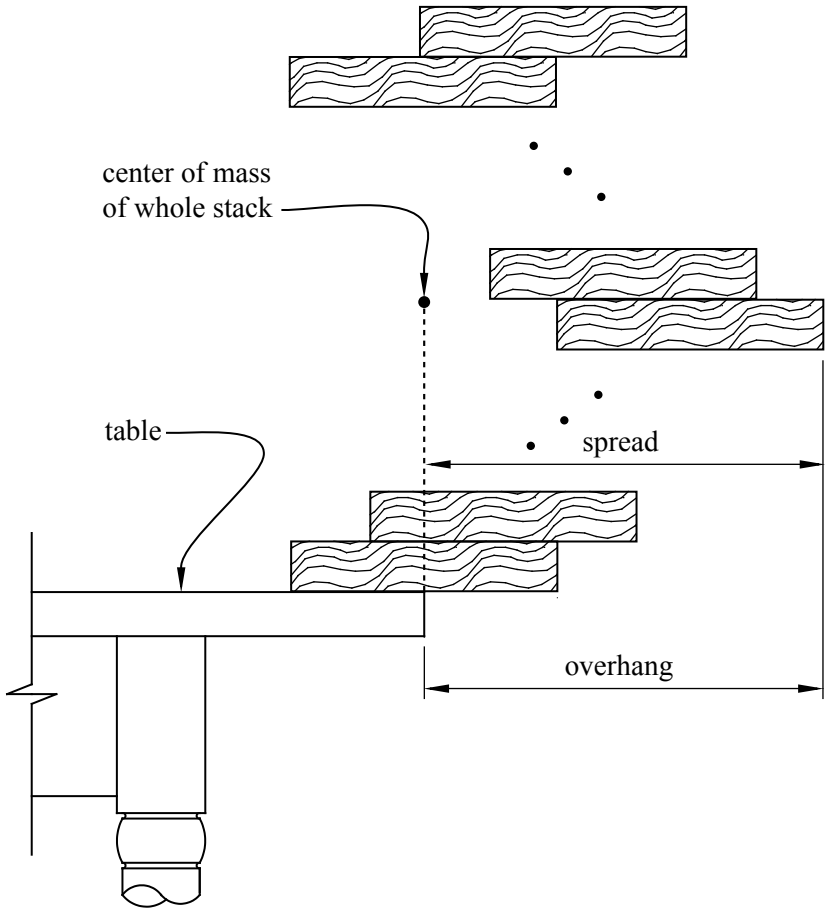



**Figure 14.6** The overhang is maximized by maximizing the spread and then placing the stack so that the center of mass is at the edge of the table.

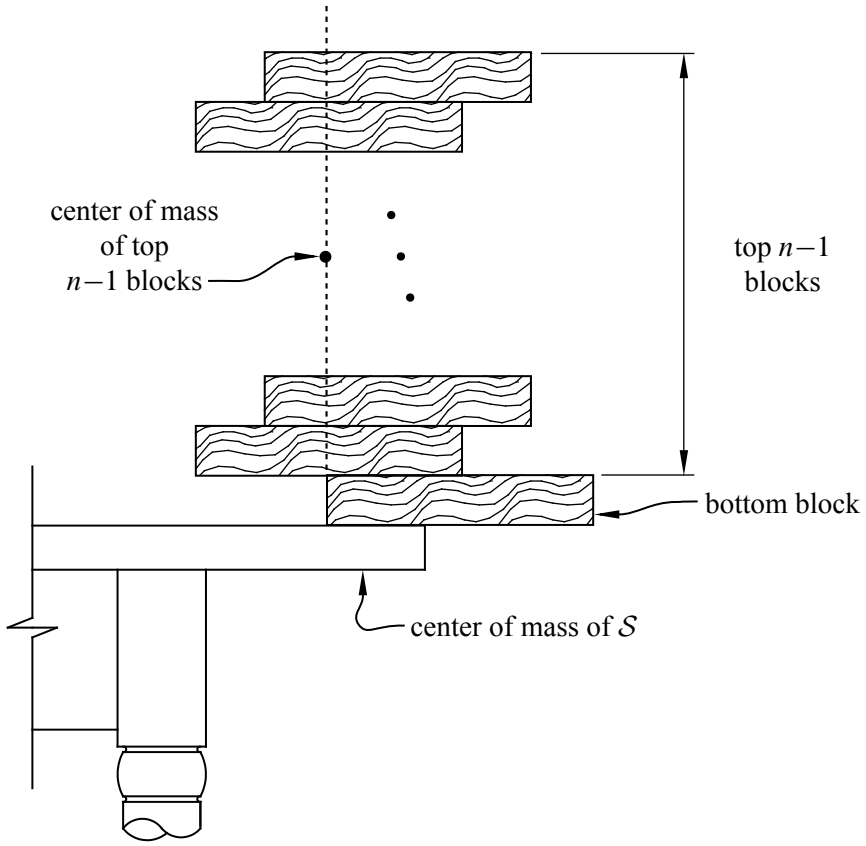



**Figure 14.7** The scenario where the bottom block is the rightmost block. In this case, the spread is maximized by having the center of mass of the top $n-1$ blocks be directly over the left edge of the bottom block.

to the left of the right edge of the bottom block and so the spread for $\mathcal{S}$ is

$$1 - \frac{1}{2n}. \tag{14.18}$$

For example, see Figure 14.7.

In fact, the scenario just described is easily achieved by arranging the blocks as shown in Figure 14.8, in which case we have the spread given by equation 14.18. For example, the spread is $3/4$ for 2 blocks, $5/6$ for 3 blocks, $7/8$ for 4 blocks, etc.

Can we do any better? The best spread in Case 1 is always less than 1, which means that we cannot get a block fully out over the edge of the table in this scenario. Maybe our intuition was right that we can't do better. Before we jump to any false conclusions, however, let's see what happens in the other case.

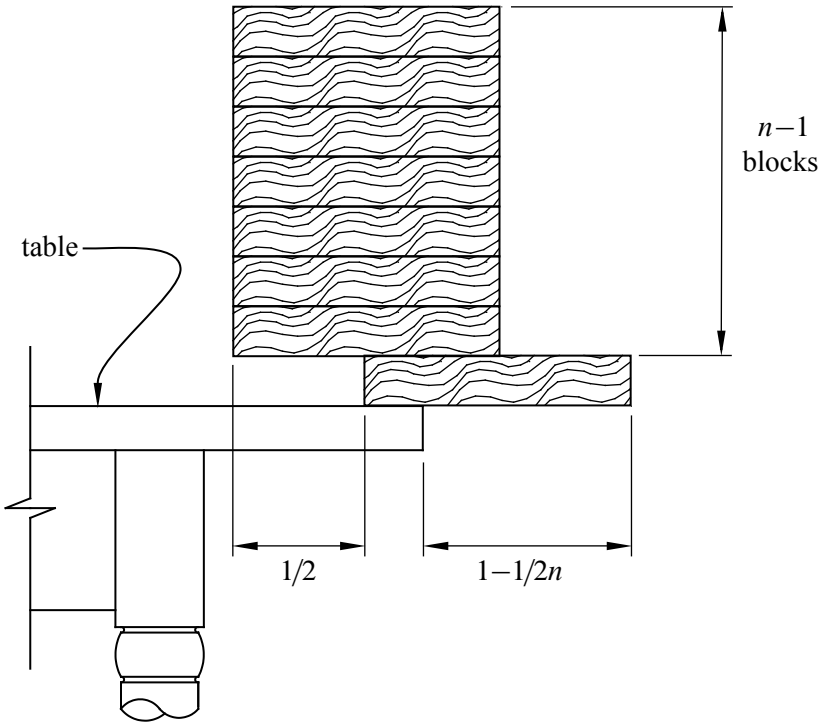



**Figure 14.8** A method for achieving spread (and hence overhang) $1 - 1/2n$ with $n$ blocks, where the bottom block is the rightmost block.

**Case 2:** *The rightmost block in $\mathcal{S}$ is among the top $n-1$ blocks.* In this case, the spread is maximized by placing the top $n-1$ blocks so that their center of mass is directly over the *right* end of the bottom block. This means that the center of mass for $\mathcal{S}$ is at location

$$\frac{(n-1)\cdot C + 1\cdot\left(C-\frac{1}{2}\right)}{n} = C - \frac{1}{2n}$$

where $C$ is the location of the center of mass of the top $n-1$ blocks. In other words, the center of mass of $\mathcal{S}$ is $1/2n$ to the left of the center of mass of the top $n-1$ blocks. (The difference is due to the effect of the bottom block, whose center of mass is $1/2$ unit to the left of $C$.) This means that the spread of $\mathcal{S}$ is $1/2n$ greater than the spread of the top $n-1$ blocks (because we are in the case where the rightmost block is among the top $n-1$ blocks.)

Since the rightmost block is among the top $n-1$ blocks, the spread for $\mathcal{S}$ is maximized by maximizing the spread for the top $n-1$ blocks. Hence the maximum spread for $\mathcal{S}$ in this case is

$$S_{n-1} + \frac{1}{2n} \tag{14.19}$$

where $S_{n-1}$ is the maximum possible spread for $n-1$ blocks (using any strategy).

We are now almost done. There are only two cases to consider when designing a stack with maximum spread and we have analyzed both of them. This means that we can combine equation 14.18 from Case 1 with equation 14.19 from Case 2 to conclude that

$$S_n = \max\left\{1 - \frac{1}{2n},\ S_{n-1} + \frac{1}{2n}\right\} \tag{14.20}$$

for any $n > 1$.

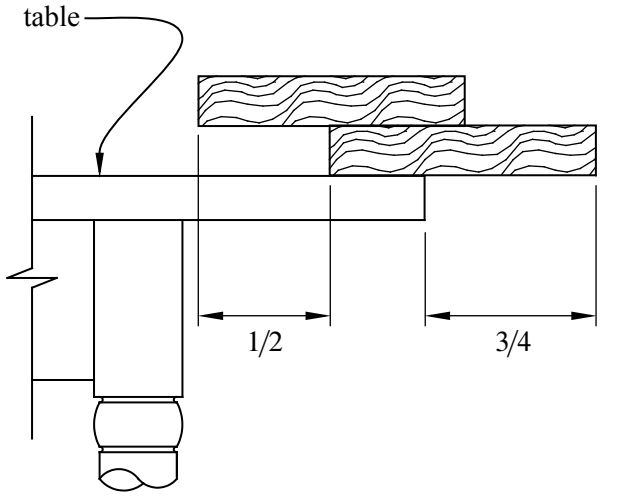
Uh-oh. This looks complicated. Maybe we are not almost done after all!

Equation 14.20 is an example of a *recurrence*. We will describe numerous techniques for solving recurrences in a later chapter, but, fortunately, equation 14.20 is simple enough that we can solve it directly.

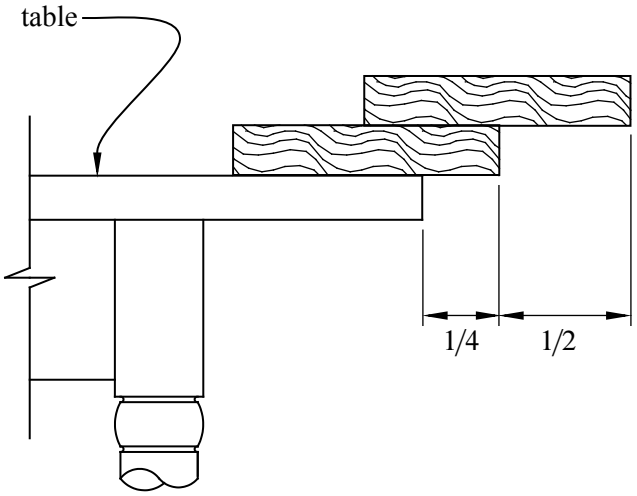
One of the first things to do when you have a recurrence is to get a feel for it by computing the first few terms. This often gives clues about a way to solve the recurrence, as it will in this case.

We already know that $S_1 = 1/2$. What about $S_2$? From equation 14.20, we find that

$$\begin{aligned} S_2 &= \max\left\{1 - \frac{1}{4},\ \frac{1}{2} + \frac{1}{4}\right\} \\ &= 3/4. \end{aligned}$$




(a)




(b)

**Figure 14.9** Two ways to achieve spread (and hence overhang) $3/4$ with $n = 2$ blocks. The first way (a) is from Case 1 and the second (b) is from Case 2.

Both cases give the same spread, albeit by different approaches. For example, see Figure 14.9.

That was easy enough. What about $S_3$?

$$\begin{aligned} S_3 &= \max\left\{1-\frac{1}{6},\ \frac{3}{4}+\frac{1}{6}\right\} \\ &= \max\left\{\frac{5}{6},\ \frac{11}{12}\right\} \\ &= \frac{11}{12}. \end{aligned}$$

As we can see, the method provided by Case 2 is the best. Let's check $n = 4$.

$$\begin{aligned} S_4 &= \max\left\{1-\frac{1}{8},\ \frac{11}{12}+\frac{1}{8}\right\} \\ &= \frac{25}{24}. \end{aligned} \tag{14.21}$$

Wow! This is a breakthrough—for two reasons. First, equation 14.21 tells us that by using only 4 blocks, we can make a stack so that one of the blocks is hanging out completely over the edge of the table. The two ways to do this are shown in Figure 14.10.

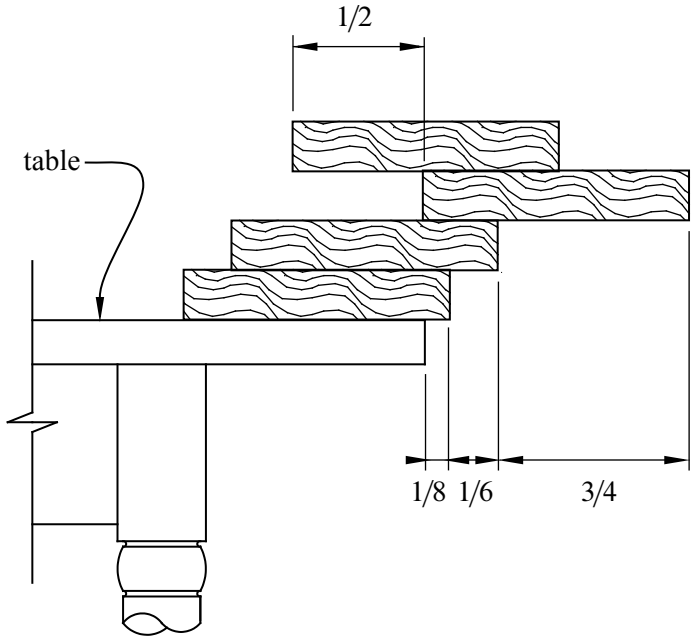
The second reason that equation 14.21 is important is that we now know that $S_4 > 1$, which means that we no longer have to worry about Case 1 for $n > 4$ since Case 1 never achieves spread greater than 1. Moreover, even for $n \leq 4$, we have now seen that the spread achieved by Case 1 never exceeds the spread achieved by Case 2, and they can be equal only for $n = 1$ and $n = 2$. This means that

$$S_n = S_{n-1} + \frac{1}{2n} \tag{14.22}$$

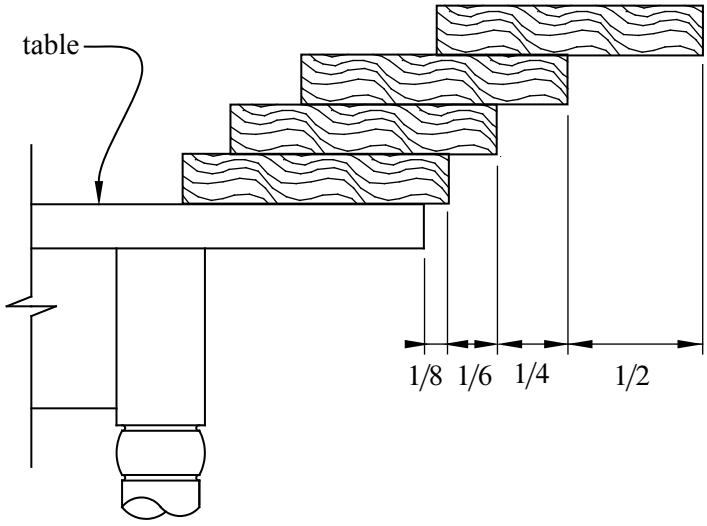
for all $n > 1$ since we have shown that the best spread can always be achieved using Case 2.

The recurrence in equation 14.22 is much easier to solve than the one we started with in equation 14.20. We can solve it by expanding the equation as follows:

$$\begin{aligned} S_n &= S_{n-1} + \frac{1}{2n} \\ &= S_{n-2} + \frac{1}{2(n-1)} + \frac{1}{2n} \\ &= S_{n-3} + \frac{1}{2(n-2)} + \frac{1}{2(n-1)} + \frac{1}{2n} \end{aligned}$$




(a)




(b)

**Figure 14.10** The two ways to achieve spread (and overhang) $25/24$. The method in (a) uses Case 1 for the top 2 blocks and Case 2 for the others. The method in (b) uses Case 2 for every block that is added to the stack.

and so on. This suggests that

$$S_n = \sum_{i=1}^{n} \frac{1}{2i}, \tag{14.23}$$

which is, indeed, the case.

Equation 14.23 can be verified by induction. The base case when $n = 1$ is true since we know that $S_1 = 1/2$. The inductive step follows from equation 14.22.

So we now know the maximum possible spread and hence the maximum possible overhang for any stable stack of books. Are we done? Not quite. Although we know that $S_4 > 1$, we still don't know how big the sum $\sum_{i=1}^{n} \frac{1}{2i}$ can get.

It turns out that $S_n$ is very close to a famous sum known as the $n$th Harmonic number $H_n$.

### 14.4.3 Harmonic Numbers

**Definition 14.4.1.** The *nth Harmonic number* is

$$H_n ::= \sum_{i=1}^{n} \frac{1}{i}.$$

So equation 14.23 means that

$$S_n = \frac{H_n}{2}. \tag{14.24}$$

The first few Harmonic numbers are easy to compute. For example,

$$H_4 = 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} = \frac{25}{12}.$$

There is good news and bad news about Harmonic numbers. The bad news is that there is no closed-form expression known for the Harmonic numbers. The good news is that we can use Theorem 14.3.2 to get close upper and lower bounds on $H_n$. In particular, since

$$\int_1^n \frac{1}{x}\, dx = \ln(x) \Big|_1^n = \ln(n),$$

Theorem 14.3.2 means that

$$\ln(n) + \frac{1}{n} \leq H_n \leq \ln(n) + 1. \tag{14.25}$$

In other words, the $n$th Harmonic number is very close to $\ln(n)$.

Because the Harmonic numbers frequently arise in practice, mathematicians have worked hard to get even better approximations for them. In fact, it is now known that

$$H_n = \ln(n) + \gamma + \frac{1}{2n} + \frac{1}{12n^2} + \frac{\epsilon(n)}{120n^4} \tag{14.26}$$

Here $\gamma$ is a value $0.577215664\ldots$ called *Euler's constant*, and $\epsilon(n)$ is between 0 and 1 for all $n$. We will not prove this formula.

We are now finally done with our analysis of the block stacking problem. Plugging the value of $H_n$ into equation 14.24, we find that the maximum overhang for $n$ blocks is very close to $\frac{1}{2}\ln(n)$. Since $\ln(n)$ grows to infinity as $n$ increases, this means that if we are given enough blocks (in theory anyway), we can get a block to hang out arbitrarily far over the edge of the table. Of course, the number of blocks we need will grow as an exponential function of the overhang, so it will probably take you a long time to achieve an overhang of 2 or 3, never mind an overhang of 100.

### 14.4.4 Asymptotic Equality

For cases like equation 14.26 where we understand the growth of a function like $H_n$ up to some (unimportant) error terms, we use a special notation, $\sim$, to denote the leading term of the function. For example, we say that $H_n \sim \ln(n)$ to indicate that the leading term of $H_n$ is $\ln(n)$. More precisely:

**Definition 14.4.2.** For functions $f, g : \mathbb{R} \to \mathbb{R}$, we say $f$ is *asymptotically equal* to $g$, in symbols,

$$f(x) \sim g(x)$$

iff

$$\lim_{x\to\infty} f(x)/g(x) = 1.$$

Although it is tempting to write $H_n \sim \ln(n) + \gamma$ to indicate the two leading terms, this is not really right. According to Definition 14.4.2, $H_n \sim \ln(n) + c$ where $c$ is *any constant*. The correct way to indicate that $\gamma$ is the second-largest term is $H_n - \ln(n) \sim \gamma$.

The reason that the $\sim$ notation is useful is that often we do not care about lower order terms. For example, if $n = 100$, then we can compute $H(n)$ to great precision using only the two leading terms:

$$|H_n - \ln(n) - \gamma| \le \left|\frac{1}{200} - \frac{1}{120000} + \frac{1}{120\cdot 100^4}\right| < \frac{1}{200}.$$

We will spend a lot more time talking about asymptotic notation at the end of the chapter. But for now, let's get back to using sums.

## 14.5 Products

We've covered several techniques for finding closed forms for sums but no methods for dealing with products. Fortunately, we do not need to develop an entirely new set of tools when we encounter a product such as

$$n! ::= \prod_{i=1}^{n} i. \tag{14.27}$$

That's because we can convert any product into a sum by taking a logarithm. For example, if

$$P = \prod_{i=1}^{n} f(i),$$

then

$$\ln(P) = \sum_{i=1}^{n} \ln(f(i)).$$

We can then apply our summing tools to find a closed form (or approximate closed form) for $\ln(P)$ and then exponentiate at the end to undo the logarithm.

For example, let's see how this works for the factorial function $n!$ We start by taking the logarithm:

$$\begin{aligned}
\ln(n!) &= \ln(1 \cdot 2 \cdot 3 \cdots (n-1) \cdot n) \\
&= \ln(1) + \ln(2) + \ln(3) + \cdots + \ln(n-1) + \ln(n) \\
&= \sum_{i=1}^{n} \ln(i).
\end{aligned}$$

Unfortunately, no closed form for this sum is known. However, we can apply Theorem 14.3.2 to find good closed-form bounds on the sum. To do this, we first compute

$$\begin{aligned}
\int_1^n \ln(x)\, dx &= x \ln(x) - x \Big|_1^n \\
&= n \ln(n) - n + 1.
\end{aligned}$$

Plugging into Theorem 14.3.2, this means that

$$n \ln(n) - n + 1 \;\le\; \sum_{i=1}^{n} \ln(i) \;\le\; n \ln(n) - n + 1 + \ln(n).$$

Exponentiating then gives

$$\frac{n^n}{e^{n-1}} \leq n! \leq \frac{n^{n+1}}{e^{n-1}}. \tag{14.28}$$

This means that $n!$ is within a factor of $n$ of $n^n/e^{n-1}$.

### 14.5.1 Stirling's Formula

$n!$ is probably the most commonly used product in discrete mathematics, and so mathematicians have put in the effort to find much better closed-form bounds on its value. The most useful bounds are given in Theorem 14.5.1.

**Theorem 14.5.1** (Stirling's Formula)**.** *For all* $n \geq 1$,

$$n! = \sqrt{2\pi n}\left(\frac{n}{e}\right)^n e^{\epsilon(n)}$$

*where*

$$\frac{1}{12n+1} \leq \epsilon(n) \leq \frac{1}{12n}.$$

Theorem 14.5.1 can be proved by induction on $n$, but the details are a bit painful (even for us) and so we will not go through them here.

There are several important things to notice about Stirling's Formula. First, $\epsilon(n)$ is always positive. This means that

$$n! > \sqrt{2\pi n}\left(\frac{n}{e}\right)^n \tag{14.29}$$

for all $n \in \mathbb{N}^+$.

Second, $\epsilon(n)$ tends to zero as $n$ gets large. This means that

$$n! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n \tag{14.30}$$

which is rather surprising. After all, who would expect both $\pi$ and $e$ to show up in a closed-form expression that is asymptotically equal to $n!$?

Third, $\epsilon(n)$ is small even for small values of $n$. This means that Stirling's Formula provides good approximations for $n!$ for most all values of $n$. For example, if we use

$$\sqrt{2\pi n}\left(\frac{n}{e}\right)^n$$

as the approximation for $n!$, as many people do, we are guaranteed to be within a factor of

$$e^{\epsilon(n)} \leq e^{\frac{1}{12n}}$$

| *Approximation* | $n \geq 1$ | $n \geq 10$ | $n \geq 100$ | $n \geq 1000$ |
|---|---|---|---|---|
| $\sqrt{2\pi n}\left(\frac{n}{e}\right)^n$ | $< 10\%$ | $< 1\%$ | $< 0.1\%$ | $< 0.01\%$ |
| $\sqrt{2\pi n}\left(\frac{n}{e}\right)^n e^{1/12n}$ | $< 1\%$ | $< 0.01\%$ | $< 0.0001\%$ | $< 0.000001\%$ |

**Table 14.1** Error bounds on common approximations for $n!$ from Theorem 14.5.1. For example, if $n \geq 100$, then $\sqrt{2\pi n}\left(\frac{n}{e}\right)^n$ approximates $n!$ to within 0.1%.

of the correct value. For $n \geq 10$, this means we will be within 1% of the correct value. For $n \geq 100$, the error will be less than 0.1%.

If we need an even closer approximation for $n!$, then we could use either

$$\sqrt{2\pi n}\left(\frac{n}{e}\right)^n e^{1/12n}$$

or

$$\sqrt{2\pi n}\left(\frac{n}{e}\right)^n e^{1/(12n+1)}$$

depending on whether we want an upper bound or a lower bound, respectively. By Theorem 14.5.1, we know that both bounds will be within a factor of

$$e^{\frac{1}{12n}-\frac{1}{12n+1}} = e^{\frac{1}{144n^2+12n}}$$

of the correct value. For $n \geq 10$, this means that either bound will be within 0.01% of the correct value. For $n \geq 100$, the error will be less than 0.0001%.

For quick future reference, these facts are summarized in Corollary 14.5.2 and Table 14.1.

**Corollary 14.5.2.**

$$n! < \sqrt{2\pi n}\left(\frac{n}{e}\right)^n \cdot \begin{cases} 1.09 & \text{for } n \geq 1, \\ 1.009 & \text{for } n \geq 10, \\ 1.0009 & \text{for } n \geq 100. \end{cases}$$

## 14.6 Double Trouble

Sometimes we have to evaluate sums of sums, otherwise known as *double summations*. This sounds hairy, and sometimes it is. But usually, it is straightforward—you just evaluate the inner sum, replace it with a closed form, and then evaluate the

outer sum (which no longer has a summation inside it). For example,[8]

$$\begin{aligned}
\sum_{n=0}^{\infty}\left(y^n\sum_{i=0}^{n}x^i\right) &= \sum_{n=0}^{\infty}\left(y^n\frac{1-x^{n+1}}{1-x}\right) && \text{equation 14.2}\\
&= \left(\frac{1}{1-x}\right)\sum_{n=0}^{\infty}y^n - \left(\frac{1}{1-x}\right)\sum_{n=0}^{\infty}y^n x^{n+1}\\
&= \frac{1}{(1-x)(1-y)} - \left(\frac{x}{1-x}\right)\sum_{n=0}^{\infty}(xy)^n && \text{Theorem 14.1.1}\\
&= \frac{1}{(1-x)(1-y)} - \frac{x}{(1-x)(1-xy)} && \text{Theorem 14.1.1}\\
&= \frac{(1-xy)-x(1-y)}{(1-x)(1-y)(1-xy)}\\
&= \frac{1-x}{(1-x)(1-y)(1-xy)}\\
&= \frac{1}{(1-y)(1-xy)}.
\end{aligned}$$

When there's no obvious closed form for the inner sum, a special trick that is often useful is to try *exchanging the order of summation.* For example, suppose we want to compute the sum of the first $n$ Harmonic numbers

$$\sum_{k=1}^{n}H_k = \sum_{k=1}^{n}\sum_{j=1}^{k}\frac{1}{j} \tag{14.31}$$

For intuition about this sum, we can apply Theorem 14.3.2 to equation 14.25 to conclude that the sum is close to

$$\int_1^n \ln(x)\,dx = x\ln(x) - x\Big|_1^n = n\ln(n) - n + 1.$$

Now let's look for an exact answer. If we think about the pairs $(k, j)$ over which

[8] Ok, so maybe this one is a little hairy, but it is also fairly straightforward. Wait till you see the next one!

we are summing, they form a triangle:

| | | $j$ | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | ... | $n$ |
| $k$ | 1 | 1 | | | | | | |
| | 2 | 1 | 1/2 | | | | | |
| | 3 | 1 | 1/2 | 1/3 | | | | |
| | 4 | 1 | 1/2 | 1/3 | 1/4 | | | |
| | | ... | | | | | | |
| | $n$ | 1 | 1/2 | | ... | | | $1/n$ |

The summation in equation 14.31 is summing each row and then adding the row sums. Instead, we can sum the columns and then add the column sums. Inspecting the table we see that this double sum can be written as

$$\begin{aligned}
\sum_{k=1}^{n} H_k &= \sum_{k=1}^{n}\sum_{j=1}^{k}\frac{1}{j} \\
&= \sum_{j=1}^{n}\sum_{k=j}^{n}\frac{1}{j} \\
&= \sum_{j=1}^{n}\frac{1}{j}\sum_{k=j}^{n}1 \\
&= \sum_{j=1}^{n}\frac{1}{j}(n-j+1) \\
&= \sum_{j=1}^{n}\frac{n+1}{j} - \sum_{j=1}^{n}\frac{j}{j} \\
&= (n+1)\sum_{j=1}^{n}\frac{1}{j} - \sum_{j=1}^{n}1 \\
&= (n+1)H_n - n. \qquad (14.32)
\end{aligned}$$

## 14.7 Asymptotic Notation

Asymptotic notation is a shorthand used to give a quick measure of the behavior of a function $f(n)$ as $n$ grows large. For example, the asymptotic notation $\sim$ of Definition 14.4.2 is a binary relation indicating that two functions grow at the *same* rate. There is also a binary relation indicating that one function grows at a significantly *slower* rate than another.

### 14.7.1 Little Oh

**Definition 14.7.1.** For functions $f, g : \mathbb{R} \to \mathbb{R}$, with $g$ nonnegative, we say $f$ is *asymptotically smaller* than $g$, in symbols,

$$f(x) = o(g(x)),$$

iff

$$\lim_{x\to\infty} f(x)/g(x) = 0.$$

For example, $1000x^{1.9} = o(x^2)$, because $1000x^{1.9}/x^2 = 1000/x^{0.1}$ and since $x^{0.1}$ goes to infinity with $x$ and 1000 is constant, we have $\lim_{x\to\infty} 1000x^{1.9}/x^2 = 0$. This argument generalizes directly to yield

**Lemma 14.7.2.** *$x^a = o(x^b)$ for all nonnegative constants $a < b$.*

Using the familiar fact that $\log x < x$ for all $x > 1$, we can prove

**Lemma 14.7.3.** *$\log x = o(x^\epsilon)$ for all $\epsilon > 0$.*

*Proof.* Choose $\epsilon > \delta > 0$ and let $x = z^\delta$ in the inequality $\log x < x$. This implies

$$\log z < z^\delta/\delta = o(z^\epsilon) \qquad \text{by Lemma 14.7.2.} \tag{14.33}$$

■

**Corollary 14.7.4.** *$x^b = o(a^x)$ for any $a, b \in \mathbb{R}$ with $a > 1$.*

Lemma 14.7.3 and Corollary 14.7.4 can also be proved using l'Hôpital's Rule or the McLaurin Series for $\log x$ and $e^x$. Proofs can be found in most calculus texts.

### 14.7.2 Big Oh

Big Oh is the most frequently used asymptotic notation. It is used to give an upper bound on the growth of a function, such as the running time of an algorithm.

**Definition 14.7.5.** Given nonnegative functions $f, g : \mathbb{R} \to \mathbb{R}$, we say that

$$f = O(g)$$

iff

$$\limsup_{x\to\infty} f(x)/g(x) < \infty.$$

This definition[9] makes it clear that

**Lemma 14.7.6.** *If $f = o(g)$ or $f \sim g$, then $f = O(g)$.*

*Proof.* $\lim f/g = 0$ or $\lim f/g = 1$ implies $\lim f/g < \infty$. ■

It is easy to see that the converse of Lemma 14.7.6 is not true. For example, $2x = O(x)$, but $2x \not\sim x$ and $2x \neq o(x)$.

The usual formulation of Big Oh spells out the definition of lim sup without mentioning it. Namely, here is an equivalent definition:

**Definition 14.7.7.** Given functions $f, g : \mathbb{R} \to \mathbb{R}$, we say that

$$f = O(g)$$

iff there exists a constant $c \geq 0$ and an $x_0$ such that for all $x \geq x_0$, $|f(x)| \leq cg(x)$.

This definition is rather complicated, but the idea is simple: $f(x) = O(g(x))$ means $f(x)$ is less than or equal to $g(x)$, except that we're willing to ignore a constant factor, namely, $c$, and to allow exceptions for small $x$, namely, $x < x_0$.

We observe,

**Lemma 14.7.8.** *If $f = o(g)$, then it is* not *true that $g = O(f)$.*

[9] We can't simply use the limit as $x \to \infty$ in the definition of $O()$, because if $f(x)/g(x)$ oscillates between, say, 3 and 5 as $x$ grows, then $f = O(g)$ because $f \leq 5g$, but $\lim_{x\to\infty} f(x)/g(x)$ does not exist. So instead of limit, we use the technical notion of lim sup. In this oscillating case, $\limsup_{x\to\infty} f(x)/g(x) = 5$.

The precise definition of lim sup is

$$\limsup_{x\to\infty} h(x) ::= \lim_{x\to\infty} \text{lub}_{y\geq x} h(y),$$

where "lub" abbreviates "least upper bound."

*Proof.*

$$\lim_{x\to\infty} \frac{g(x)}{f(x)} = \frac{1}{\lim_{x\to\infty} f(x)/g(x)} = \frac{1}{0} = \infty,$$

so $g \neq O(f)$.

■

**Proposition 14.7.9.** $100x^2 = O(x^2)$.

*Proof.* Choose $c = 100$ and $x_0 = 1$. Then the proposition holds, since for all $x \geq 1$, $\left|100x^2\right| \leq 100x^2$. ■

**Proposition 14.7.10.** $x^2 + 100x + 10 = O(x^2)$.

*Proof.* $(x^2+100x+10)/x^2 = 1+100/x+10/x^2$ and so its limit as $x$ approaches infinity is $1+0+0 = 1$. So in fact, $x^2+100x+10 \sim x^2$, and therefore $x^2+100x+10 = O(x^2)$. Indeed, it's conversely true that $x^2 = O(x^2 + 100x + 10)$. ■

Proposition 14.7.10 generalizes to an arbitrary polynomial:

**Proposition 14.7.11.** $a_k x^k + a_{k-1}x^{k-1} + \cdots + a_1 x + a_0 = O(x^k)$.

We'll omit the routine proof.

Big Oh notation is especially useful when describing the running time of an algorithm. For example, the usual algorithm for multiplying $n \times n$ matrices uses a number of operations proportional to $n^3$ in the worst case. This fact can be expressed concisely by saying that the running time is $O(n^3)$. So this asymptotic notation allows the speed of the algorithm to be discussed without reference to constant factors or lower-order terms that might be machine specific. It turns out that there is another, ingenious matrix multiplication procedure that uses $O(n^{2.55})$ operations. This procedure will therefore be much more efficient on large enough matrices. Unfortunately, the $O(n^{2.55})$-operation multiplication procedure is almost never used in practice because it happens to be less efficient than the usual $O(n^3)$ procedure on matrices of practical size.[10]

### 14.7.3 Theta

Sometimes we want to specify that a running time $T(n)$ is precisely quadratic up to constant factors (both upper bound *and* lower bound). We could do this by saying that $T(n) = O(n^2)$ and $n^2 = O(T(n))$, but rather than say both, mathematicians have devised yet another symbol, $\Theta$, to do the job.

[10] It is even conceivable that there is an $O(n^2)$ matrix multiplication procedure, but none is known.

**Definition 14.7.12.**

$$f = \Theta(g) \quad \text{iff} \quad f = O(g) \text{ and } g = O(f).$$

The statement $f = \Theta(g)$ can be paraphrased intuitively as "$f$ and $g$ are equal to within a constant factor."

The Theta notation allows us to highlight growth rates and allow suppression of distracting factors and low-order terms. For example, if the running time of an algorithm is

$$T(n) = 10n^3 - 20n^2 + 1,$$

then we can more simply write

$$T(n) = \Theta(n^3).$$

In this case, we would say that $T$ *is of order* $n^3$ or that $T(n)$ *grows cubically*, which is probably what we really want to know. Another such example is

$$\pi^2 3^{x-7} + \frac{(2.7x^{113} + x^9 - 86)^4}{\sqrt{x}} - 1.08^{3x} = \Theta(3^x).$$

Just knowing that the running time of an algorithm is $\Theta(n^3)$, for example, is useful, because if $n$ doubles we can predict that the running time will *by and large*[11] increase by a factor of at most 8 for large $n$. In this way, Theta notation preserves information about the scalability of an algorithm or system. Scalability is, of course, a big issue in the design of algorithms and systems.

### 14.7.4 Pitfalls with Asymptotic Notation

There is a long list of ways to make mistakes with asymptotic notation. This section presents some of the ways that Big Oh notation can lead to ruin and despair. With minimal effort, you can cause just as much chaos with the other symbols.

#### The Exponential Fiasco

Sometimes relationships involving Big Oh are not so obvious. For example, one might guess that $4^x = O(2^x)$ since 4 is only a constant factor larger than 2. This reasoning is incorrect, however; $4^x$ actually grows as the square of $2^x$.

[11] Since $\Theta(n^3)$ only implies that the running time, $T(n)$, is between $cn^3$ and $dn^3$ for constants $0 < c < d$, the time $T(2n)$ could regularly exceed $T(n)$ by a factor as large as $8d/c$. The factor is sure to be close to 8 for all large $n$ only if $T(n) \sim n^3$.

### Constant Confusion

Every constant is $O(1)$. For example, $17 = O(1)$. This is true because if we let $f(x) = 17$ and $g(x) = 1$, then there exists a $c > 0$ and an $x_0$ such that $|f(x)| \leq cg(x)$. In particular, we could choose $c = 17$ and $x_0 = 1$, since $|17| \leq 17 \cdot 1$ for all $x \geq 1$. We can construct a false theorem that exploits this fact.

**False Theorem 14.7.13.**

$$\sum_{i=1}^{n} i = O(n)$$

*Bogus proof.* Define $f(n) = \sum_{i=1}^{n} i = 1+2+3+\cdots+n$. Since we have shown that every constant $i$ is $O(1)$, $f(n) = O(1) + O(1) + \cdots + O(1) = O(n)$. ■

Of course in reality $\sum_{i=1}^{n} i = n(n+1)/2 \neq O(n)$.

The error stems from confusion over what is meant in the statement $i = O(1)$. For any *constant* $i \in \mathbb{N}$ it is true that $i = O(1)$. More precisely, if $f$ is any constant function, then $f = O(1)$. But in this False Theorem, $i$ is not constant —it ranges over a set of values $0, 1, \ldots, n$ that depends on $n$.

And anyway, we should not be adding $O(1)$'s as though they were numbers. We never even defined what $O(g)$ means by itself; it should only be used in the context "$f = O(g)$" to describe a relation between functions $f$ and $g$.

### Lower Bound Blunder

Sometimes people incorrectly use Big Oh in the context of a lower bound. For example, they might say, "The running time, $T(n)$, is at least $O(n^2)$," when they probably mean "$n^2 = O(T(n))$." [12]

### Equality Blunder

The notation $f = O(g)$ is too firmly entrenched to avoid, but the use of "=" is really regrettable. For example, if $f = O(g)$, it seems quite reasonable to write $O(g) = f$. But doing so might tempt us to the following blunder: because $2n = O(n)$, we can say $O(n) = 2n$. But $n = O(n)$, so we conclude that $n = O(n) = 2n$, and therefore $n = 2n$. To avoid such nonsense, we will never write "$O(f) = g$."

Similarly, you will often see statements like

$$H_n = \ln(n) + \gamma + O\left(\frac{1}{n}\right)$$

[12] This would more usually be expressed as "$T(n) = \Omega(n^2)$."

or

$$n! = (1 + o(1))\sqrt{2\pi n}\left(\frac{n}{e}\right)^n.$$

In such cases, the true meaning is

$$H_n = \ln(n) + \gamma + f(n)$$

for some $f(n)$ where $f(n) = O(1/n)$, and

$$n! = (1 + g(n))\sqrt{2\pi n}\left(\frac{n}{e}\right)^n$$

where $g(n) = o(1)$. These last transgressions are OK as long as you (and your reader) know what you mean.

### 14.7.5 Omega

Suppose you want to make a statement of the form "the running time of the algorithm is a least…." Can you say it is "at least $O(n^2)$"? No! This statement is meaningless since big-oh can only be used for *upper* bounds. For lower bounds, we use a different symbol, called "big-Omega."

**Definition 14.7.14.** Given functions $f, g : \mathbb{R} \to \mathbb{R}$, define

$$f = \Omega(g)$$

to mean

$$g = O(f).$$

For example, $x^2 = \Omega(x)$, $2^x = \Omega(x^2)$, and $x/100 = \Omega(100x + \sqrt{x})$.

So if the running time of your algorithm on inputs of size $n$ is $T(n)$, and you want to say it is at least quadratic, say

$$T(n) = \Omega(n^2).$$

Likewise, there is also a symbol called little-omega, analogous to little-oh, to denote that one function grows strictly faster than another function.

**Definition 14.7.15.** For functions $f, g : \mathbb{R} \to \mathbb{R}$ with $f$ nonnegative, define

$$f = \omega(g)$$

to mean

$$g = o(f).$$

For example, $x^{1.5} = \omega(x)$ and $\sqrt{x} = \omega(\ln^2(x))$.

The little-omega symbol is not as widely used as the other asymptotic symbols we defined.

## Problems for Section 14.1

### Class Problems

**Problem 14.1.**
We begin with two large glasses. The first glass contains a pint of water, and the second contains a pint of wine. We pour 1/3 of a pint from the first glass into the second, stir up the wine/water mixture in the second glass, and then pour 1/3 of a pint of the mix back into the first glass and repeat this pouring back-and-forth process a total of $n$ times.

**(a)** Describe a closed form formula for the amount of wine in the first glass after $n$ back-and-forth pourings.

**(b)** What is the limit of the amount of wine in each glass as $n$ approaches infinity?

**Problem 14.2.**
You've seen this neat trick for evaluating a geometric sum:

$$\begin{aligned} S &= 1 + z + z^2 + \ldots + z^n \\ zS &= z + z^2 + \ldots + z^n + z^{n+1} \\ S - zS &= 1 - z^{n+1} \\ S &= \frac{1 - z^{n+1}}{1 - z} \end{aligned}$$

Use the same approach to find a closed-form expression for this sum:

$$T = 1z + 2z^2 + 3z^3 + \ldots + nz^n$$

### Homework Problems

**Problem 14.3.**
Is a Harvard degree really worth more than an MIT degree?! Let us say that a person with a Harvard degree starts with $40,000 and gets a $20,000 raise every year after graduation, whereas a person with an MIT degree starts with $30,000, but gets a 20% raise every year. Assume inflation is a fixed 8% every year. That is, $1.08 a year from now is worth $1.00 today.

**(a)** How much is a Harvard degree worth today if the holder will work for $n$ years following graduation?

**(b)** How much is an MIT degree worth in this case?

**(c)** If you plan to retire after twenty years, which degree would be worth more?

**Problem 14.4.**
Suppose you deposit \$100 into your MIT Credit Union account today, \$99 in one month from now, \$98 in two months from now, and so on. Given that the interest rate is constantly 0.3% per month, how long will it take to save \$5,000?

## Problems for Section 14.3

### Practice Problems

**Problem 14.5.**
Let

$$S ::= \sum_{n=1}^{5} n^{\frac{1}{3}}$$

Using the **Integral Method**, we can find integers, $a$, $b$, $c$, $d$, and a real number, $e$, such that

$$\int_a^b x^e \, dx \leq S \leq \int_c^d x^e \, dx$$

What are appropriate values for $a$–$e$ ?

### Exam Problems

**Problem 14.6.**
Assume $n$ is an integer larger than 1. Circle all the correct inequalities below.

Explanations are not required, but partial credit for wrong answers will not be given without them. *Hint:* You may find the graphs in Figure 14.11 helpful.

- $\sum_{i=1}^{n} \ln(i+1) \leq \ln 2 + \int_1^n \ln(x+1)dx$
- $\sum_{i=1}^{n} \ln(i+1) \leq \int_0^n \ln(x+2)dx$
- $\sum_{i=1}^{n} \frac{1}{i} \geq \int_0^n \frac{1}{x+1}dx$

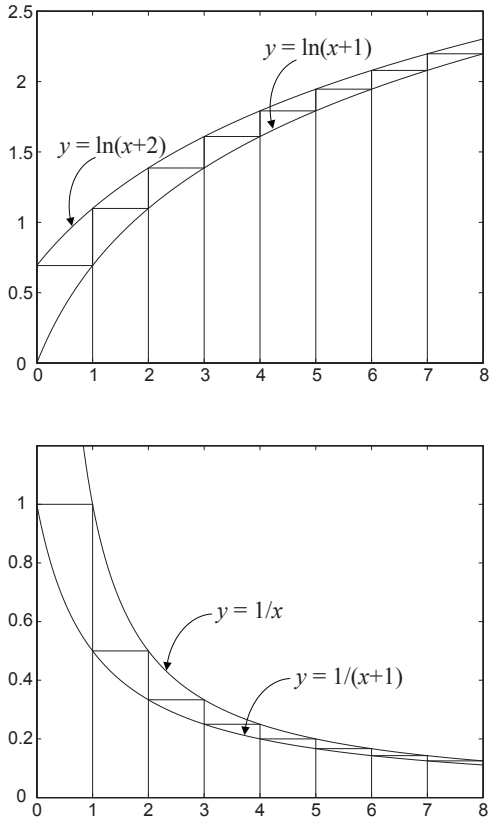



**Figure 14.11** Integral bounds for two sums

### Homework Problems

**Problem 14.7.**
Let $f : \mathbb{R}^+ \to \mathbb{R}^+$ be a weakly decreasing function. Define

$$S ::= \sum_{i=1}^{n} f(i)$$

and

$$I ::= \int_1^n f(x)\, dx.$$

Prove that

$$I + f(n) \leq S \leq I + f(1).$$

**Problem 14.8.**
Use integration to find upper and lower bounds that differ by at most 0.1 for the following sum. (You may need to add the first few terms explicitly and then use integrals to bound the sum of the remaining terms.)

$$\sum_{i=1}^{\infty} \frac{1}{(2i+1)^2}$$

## Problems for Section 14.4

### Class Problems

**Problem 14.9.**
An explorer is trying to reach the Holy Grail, which she believes is located in a desert shrine $d$ days walk from the nearest oasis. In the desert heat, the explorer must drink continuously. She can carry at most 1 gallon of water, which is enough for 1 day. However, she is free to make multiple trips carrying up to a gallon each time to create water caches out in the desert.

For example, if the shrine were $2/3$ of a day's walk into the desert, then she could recover the Holy Grail after two days using the following strategy. She leaves the oasis with 1 gallon of water, travels $1/3$ day into the desert, caches $1/3$ gallon, and then walks back to the oasis—arriving just as her water supply runs out. Then she picks up another gallon of water at the oasis, walks $1/3$ day into the desert, tops off her water supply by taking the $1/3$ gallon in her cache, walks the remaining $1/3$ day to the shrine, grabs the Holy Grail, and then walks for $2/3$ of a day back to the oasis—again arriving with no water to spare.

But what if the shrine were located farther away?

**(a)** What is the most distant point that the explorer can reach and then return to the oasis if she takes a total of only 1 gallon from the oasis?

**(b)** What is the most distant point the explorer can reach and still return to the oasis if she takes a total of only 2 gallons from the oasis? No proof is required; just do the best you can.

**(c)** The explorer will travel using a recursive strategy to go far into the desert and back drawing a total of $n$ gallons of water from the oasis. Her strategy is to build up a cache of $n-1$ gallons, plus enough to get home, a certain fraction of a day's distance into the desert. On the last delivery to the cache, instead of returning home, she proceeds recursively with her $n-1$ gallon strategy to go farther into the desert and return to the cache. At this point, the cache has just enough water left to get her home.

Prove that with $n$ gallons of water, this strategy will get her $H_n/2$ days into the desert and back, where $H_n$ is the $n$th Harmonic number:

$$H_n ::= \frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}.$$

Conclude that she can reach the shrine, however far it is from the oasis.

**(d)** Suppose that the shrine is $d = 10$ days walk into the desert. Use the asymptotic approximation $H_n \sim \ln n$ to show that it will take more than a million years for the explorer to recover the Holy Grail.

**Problem 14.10.**
There is a number $a$ such that $\sum_{i=1}^{\infty} i^p$ converges iff $p < a$. What is the value of $a$?

*Hint:* Find a value for $a$ you think that works, then apply the integral bound.

**Homework Problems**

**Problem 14.11.**
There is a bug on the edge of a 1-meter rug. The bug wants to cross to the other side of the rug. It crawls at 1 cm per second. However, at the end of each second, a malicious first-grader named Mildred Anderson *stretches* the rug by 1 meter. Assume that her action is instantaneous and the rug stretches uniformly. Thus, here's what happens in the first few seconds:

- The bug walks 1 cm in the first second, so 99 cm remain ahead.

- Mildred stretches the rug by 1 meter, which doubles its length. So now there are 2 cm behind the bug and 198 cm ahead.
- The bug walks another 1 cm in the next second, leaving 3 cm behind and 197 cm ahead.
- Then Mildred strikes, stretching the rug from 2 meters to 3 meters. So there are now $3 \cdot (3/2) = 4.5$ cm behind the bug and $197 \cdot (3/2) = 295.5$ cm ahead.
- The bug walks another 1 cm in the third second, and so on.

Your job is to determine this poor bug's fate.

**(a)** During second $i$, what *fraction* of the rug does the bug cross?

**(b)** Over the first $n$ seconds, what fraction of the rug does the bug cross altogether? Express your answer in terms of the Harmonic number $H_n$.

**(c)** The known universe is thought to be about $3 \cdot 10^{10}$ light years in diameter. How many universe diameters must the bug travel to get to the end of the rug? (This distance is NOT the inflated distance caused by the stretching but only the actual walking done by the bug).

## Problems for Section 14.7

### Practice Problems

**Problem 14.12.**
Find the least nonnegative integer, $n$, such that $f(x)$ is $O(x^n)$ when $f$ is defined by each of the expressions below.

**(a)** $2x^3 + (\log x)x^2$

**(b)** $2x^2 + (\log x)x^3$

**(c)** $(1.1)^x$

**(d)** $(0.1)^x$

**(e)** $(x^4 + x^2 + 1)/(x^3 + 1)$

**(f)** $(x^4 + 5\log x)/(x^4 + 1)$

**(g)** $2^{(3\log_2 x^2)}$

**Problem 14.13.**
Let $f(n) = n^3$. For each function $g(n)$ in the table below, indicate which of the indicated asymptotic relations hold.

| $g(n)$ | $f = O(g)$ | $f = o(g)$ | $g = O(f)$ | $g = o(f)$ |
|---|---|---|---|---|
| $6 - 5n - 4n^2 + 3n^3$ | | | | |
| $n^3 \log n$ | | | | |
| $(\sin(\pi n/2) + 2)\, n^3$ | | | | |
| $n^{\sin(\pi n/2)+2}$ | | | | |
| $\log n!$ | | | | |
| $e^{0.2n} - 100n^3$ | | | | |

**Problem 14.14.**
Circle each of the true statements below.

Explanations are not required, but partial credit for wrong answers will not be given without them.

- $n^2 \sim n^2 + n$
- $3^n = O\left(2^n\right)$
- $n^{\sin(n\pi/2)+1} = o\left(n^2\right)$
- $n = \Theta\left(\dfrac{3n^3}{(n+1)(n-1)}\right)$

**Problem 14.15.**
Show that

$$\ln(n^2!) = \Theta(n^2 \ln n)$$

**Problem 14.16.**
The quantity

$$\frac{(2n)!}{2^{2n}(n!)^2} \tag{14.34}$$

will come up later in the course (it is the probability that in $2^{2n}$ flips of a fair coin, exactly $n$ will be Heads). Show that it is asymptotically equal to $\frac{1}{\sqrt{\pi n}}$.

## Homework Problems

**Problem 14.17. (a)** Prove that $\log x < x$ for all $x > 1$ (requires elementary calculus).

**(b)** Prove that the relation, $R$, on functions such that $f \; R \; g$ iff $f = o(g)$ is a strict partial order.

**(c)** Prove that $f \sim g$ iff $f = g + h$ for some function $h = o(g)$.

**Problem 14.18.**
Indicate which of the following holds for each pair of functions $(f(n), g(n))$ in the table below. Assume $k \geq 1$, $\epsilon > 0$, and $c > 1$ are constants. Pick the four table entries you consider to be the most challenging or interesting and justify your answers to these.

| $f(n)$ | $g(n)$ | $f = O(g)$ | $f = o(g)$ | $g = O(f)$ | $g = o(f)$ | $f = \Theta(g)$ | $f \sim g$ |
|---|---|---|---|---|---|---|---|
| $2^n$ | $2^{n/2}$ | | | | | | |
| $\sqrt{n}$ | $n^{\sin(n\pi/2)}$ | | | | | | |
| $\log(n!)$ | $\log(n^n)$ | | | | | | |
| $n^k$ | $c^n$ | | | | | | |
| $\log^k n$ | $n^\epsilon$ | | | | | | |

**Problem 14.19.**
Let $f$, $g$ be nonnegative real-valued functions such that $\lim_{x\to\infty} f(x) = \infty$ and $f \sim g$.

**(a)** Give an example of $f, g$ such that NOT$(2^f \sim 2^g)$.

**(b)** Prove that $\log f \sim \log g$.

**(c)** Use Stirling's formula to prove that in fact

$$\log(n!) \sim n \log n$$

**Problem 14.20.**
Determine which of these choices

$$\Theta(n), \quad \Theta(n^2 \log n), \quad \Theta(n^2), \quad \Theta(1), \quad \Theta(2^n), \quad \Theta(2^{n \ln n}), \quad \text{none of these}$$

describes each function's asymptotic behavior. Full proofs are not required, but briefly explain your answers.

**(a)**

$$n + \ln n + (\ln n)^2$$

**(b)**

$$\frac{n^2 + 2n - 3}{n^2 - 7}$$

**(c)**

$$\sum_{i=0}^{n} 2^{2i+1}$$

**(d)**

$$\ln(n^2!)$$

**(e)**

$$\sum_{k=1}^{n} k\left(1 - \frac{1}{2^k}\right)$$

**Problem 14.21. (a)** Either prove or disprove each of the following statements.

- $n! = O((n+1)!)$
- $(n+1)! = O(n!)$
- $n! = \Theta((n+1)!)$
- $n! = o((n+1)!)$
- $(n+1)! = o(n!)$

**(b)** Show that $\left(\frac{n}{3}\right)^{n+e} = o(n!)$.

**Problem 14.22.**
Prove that $\sum_{k=1}^{n} k^6 = \Theta(n^7)$.

## Class Problems

**Problem 14.23.**
Give an elementary proof (without appealing to Stirling's formula) that $\log(n!) = \Theta(n \log n)$.

**Problem 14.24.**
Suppose $f, g : \mathbb{N}^+ \to \mathbb{N}^+$ and $f \sim g$.

**(a)** Prove that $2f \sim 2g$.

**(b)** Prove that $f^2 \sim g^2$.

**(c)** Give examples of $f$ and $g$ such that $2^f \not\sim 2^g$.

**Problem 14.25.**
Recall that for functions $f, g$ on $\mathbb{N}$, $f = O(g)$ iff

$$\exists c \in \mathbb{N}\, \exists n_0 \in \mathbb{N}\, \forall n \geq n_0 \quad c \cdot g(n) \geq |f(n)|. \tag{14.35}$$

For each pair of functions below, determine whether $f = O(g)$ and whether $g = O(f)$. In cases where one function is O() of the other, indicate the *smallest nonnegative integer*, $c$, and for that smallest $c$, the *smallest corresponding nonnegative integer* $n_0$ ensuring that condition (14.35) applies.

**(a)** $f(n) = n^2, g(n) = 3n$.

| | | | |
|---|---|---|---|
| $f = O(g)$ | YES | NO | If YES, $c =$ _____, $n_0 =$ _____ |
| $g = O(f)$ | YES | NO | If YES, $c =$ _____, $n_0 =$ _____ |

**(b)** $f(n) = (3n - 7)/(n + 4), g(n) = 4$

| | | | |
|---|---|---|---|
| $f = O(g)$ | YES | NO | If YES, $c =$ _____, $n_0 =$ _____ |
| $g = O(f)$ | YES | NO | If YES, $c =$ _____, $n_0 =$ _____ |

**(c)** $f(n) = 1 + (n \sin(n\pi/2))^2, g(n) = 3n$

| | | | |
|---|---|---|---|
| $f = O(g)$ | YES | NO | If yes, $c =$ _____ $n_0 =$ _____ |
| $g = O(f)$ | YES | NO | If yes, $c =$ _____ $n_0 =$ _____ |

**Problem 14.26.**

**False Claim.**

$$2^n = O(1). \tag{14.36}$$

Explain why the claim is false. Then identify and explain the mistake in the following bogus proof.

*Bogus proof.* The proof by induction on $n$ where the induction hypothesis, $P(n)$, is the assertion (14.36).

**base case:** $P(0)$ holds trivially.

**inductive step:** We may assume $P(n)$, so there is a constant $c > 0$ such that $2^n \leq c \cdot 1$. Therefore,

$$2^{n+1} = 2 \cdot 2^n \leq (2c) \cdot 1,$$

which implies that $2^{n+1} = O(1)$. That is, $P(n+1)$ holds, which completes the proof of the inductive step.

We conclude by induction that $2^n = O(1)$ for all $n$. That is, the exponential function is bounded by a constant.

■

**Problem 14.27. (a)** Prove that the relation, $R$, on functions such that $f \; R \; g$ iff $f = o(g)$ is a strict partial order.

**(b)** Describe two functions $f, g$ that are incomparable under big Oh:

$$f \neq O(g) \text{ AND } g \neq O(f).$$

Conclude that $R$ is not a path-total order.

**Exam Problems**

**Problem 14.28. (a)** Show that

$$(an)^{b/n} \sim 1.$$

where $a, b$ are positive constants and $\sim$ denotes asymptotic equality. *Hint:* $an = a2^{\log_2 n}$.

**(b)** You may assume that if $f(n) \geq 1$ and $g(n) \geq 1$ for all $n$, then $f \sim g \longrightarrow f^{\frac{1}{n}} \sim g^{\frac{1}{n}}$. Show that

$$\sqrt[n]{n!} = \Theta(n).$$

**Problem 14.29.**

**(a)** Define a function $f(n)$ such that $f = \Theta(n^2)$ and NOT($f \sim n^2$).

**(b)** Define a function $g(n)$ such that $g = O(n^2)$, $g \neq \Theta(n^2)$ and $g \neq o(n^2)$.

**Problem 14.30. (a)** Show that

$$(an)^{b/n} \sim 1.$$

where $a, b$ are positive constants and $\sim$ denotes asymptotic equality. *Hint:* $an = a2^{\log_2 n}$.

**(b)** Show that

$$\sqrt[n]{n!} = \Theta(n).$$

**Problem 14.31.**

**(a)** Indicate which of the following asymptotic relations below on the set of nonnegative real-valued functions are *equivalence relations*, (**E**), strict partial orders (**S**), weak partial orders (**W**), or *none* of the above (**N**).

- $f \sim g$, the "asymptotically Equal" relation.
- $f = o(g)$, the "little Oh" relation.
- $f = O(g)$, the "big Oh" relation.
- $f = \Theta(g)$, the "Theta" relation.
- $f = O(g)$ AND NOT($g = O(f)$).

**(b)** Define two functions $f, g$ that are incomparable under big Oh:

$$f \neq O(g) \text{ AND } g \neq O(f).$$

**Problem 14.32.**
Recall that if $f$ and $g$ are nonnegative real-valued functions on $\mathbb{Z}^+$, then $f = O(g)$ iff there exist $c, n_0 \in \mathbb{Z}^+$ such that

$$\forall n \geq n_0.\ f(n) \leq cg(n).$$

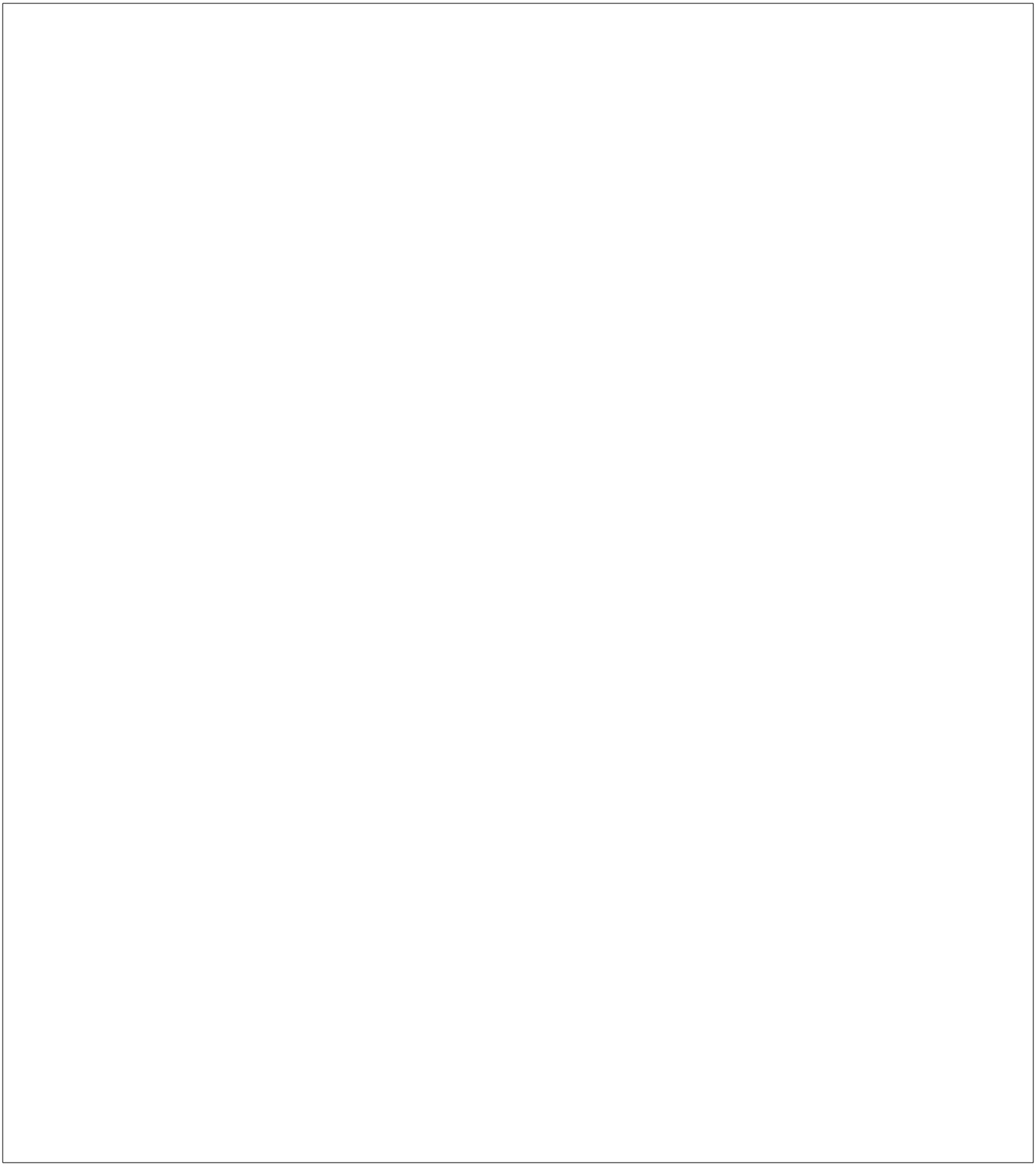
For each pair of functions $f$ and $g$ below, indicate the **smallest** $c \in \mathbb{Z}^+$, and for that smallest $c$, the **smallest corresponding** $n_0 \in \mathbb{Z}^+$, that would establish $f = O(g)$ by the definition given above. If there is no such $c$, write $\infty$.

**(a)** $f(n) = \frac{1}{2}\ln n^2, g(n) = n.$ $c =$ _____, $n_0 =$ _____

**(b)** $f(n) = n, g(n) = n \ln n.$ $c =$ _____, $n_0 =$ _____

**(c)** $f(n) = 2^n, g(n) = n^4 \ln n$ $c =$ _____, $n_0 =$ _____

**(d)** $f(n) = 3\sin\left(\frac{\pi(n-1)}{100}\right) + 2, g(n) = 0.2.$ $c =$ _____, $n_0 =$ _____



# 15 Cardinality Rules

## 15.1 Counting One Thing by Counting Another

How do you count the number of people in a crowded room? You could count heads, since for each person there is exactly one head. Alternatively, you could count ears and divide by two. Of course, you might have to adjust the calculation if someone lost an ear in a pirate raid or someone was born with three ears. The point here is that you can often *count one thing by counting another*, though some fudge factors may be required. This is a central theme of counting, from the easiest problems to the hardest. In fact, we've already seen this technique used in Theorem 5.1.5 where the number of subsets of an $n$-element set was proved to be the same as the number of length-$n$ bit-strings by describing a bijection between the subsets and the bit-strings.

The most direct way to count one thing by counting another is to find a bijection between them, since if there is a bijection between two sets, then the sets have the same size. This important fact is commonly known as the *Bijection Rule*. We've already seen it as the Mapping Rules bijective case (5.3).

### 15.1.1 The Bijection Rule

The Bijection Rule acts as a magnifier of counting ability; if you figure out the size of one set, then you can immediately determine the sizes of many other sets via bijections. For example, let's look at the two sets mentioned at the beginning of Part III:

$A$ = all ways to select a dozen doughnuts when five varieties are available

$B$ = all 16-bit sequences with exactly 4 ones

An example of an element of set $A$ is:

$$\underbrace{0\,0}_{\text{chocolate}} \quad \underbrace{\phantom{00}}_{\text{lemon-filled}} \quad \underbrace{0\,0\,0\,0\,0\,0}_{\text{sugar}} \quad \underbrace{0\,0}_{\text{glazed}} \quad \underbrace{0\,0}_{\text{plain}}$$

Here, we've depicted each doughnut with a 0 and left a gap between the different varieties. Thus, the selection above contains two chocolate doughnuts, no lemon-filled, six sugar, two glazed, and two plain. Now let's put a 1 into each of the four

gaps:

$$\underbrace{0\,0}_{\text{chocolate}}\quad 1\quad \underbrace{\phantom{0\,0}}_{\text{lemon-filled}}\quad 1\quad \underbrace{0\,0\,0\,0\,0\,0}_{\text{sugar}}\quad 1\quad \underbrace{0\,0}_{\text{glazed}}\quad 1\quad \underbrace{0\,0}_{\text{plain}}$$

and close up the gaps:

$$0011000000100100\,.$$

We've just formed a 16-bit number with exactly 4 ones —an element of $B$!

This example suggests a bijection from set $A$ to set $B$: map a dozen doughnuts consisting of:

$$c \text{ chocolate, } l \text{ lemon-filled, } s \text{ sugar, } g \text{ glazed, and } p \text{ plain}$$

to the sequence:

$$\underbrace{0\ldots0}_{c}\quad 1\quad \underbrace{0\ldots0}_{l}\quad 1\quad \underbrace{0\ldots0}_{s}\quad 1\quad \underbrace{0\ldots0}_{g}\quad 1\quad \underbrace{0\ldots0}_{p}$$

The resulting sequence always has 16 bits and exactly 4 ones, and thus is an element of $B$. Moreover, the mapping is a bijection; every such bit sequence comes from exactly one order of a dozen doughnuts. Therefore, $|A| = |B|$ by the Bijection Rule!

This example demonstrates the magnifying power of the bijection rule. We managed to prove that two very different sets are actually the same size —even though we don't know exactly how big either one is. But as soon as we figure out the size of one set, we'll immediately know the size of the other.

This particular bijection might seem frighteningly ingenious if you've not seen it before. But you'll use essentially this same argument over and over, and soon you'll consider it routine.

## 15.2 Counting Sequences

The Bijection Rule lets us count one thing by counting another. This suggests a general strategy: get really good at counting just a *few* things and then use bijections to count *everything else*. This is the strategy we'll follow. In particular, we'll get really good at counting *sequences*. When we want to determine the size of some other set $T$, we'll find a bijection from $T$ to a set of sequences $S$. Then we'll use our super-ninja sequence-counting skills to determine $|S|$, which immediately gives us $|T|$. We'll need to hone this idea somewhat as we go along, but that's pretty much the plan!

### 15.2.1 The Product Rule

The *Product Rule* gives the size of a product of sets. Recall that if $P_1, P_2, \ldots, P_n$ are sets, then

$$P_1 \times P_2 \times \ldots \times P_n$$

is the set of all sequences whose first term is drawn from $P_1$, second term is drawn from $P_2$ and so forth.

**Rule 15.2.1** (Product Rule). *If $P_1, P_2, \ldots P_n$ are finite sets, then:*

$$|P_1 \times P_2 \times \ldots \times P_n| = |P_1| \cdot |P_2| \cdots |P_n|$$

For example, suppose a *daily diet* consists of a breakfast selected from set $B$, a lunch from set $L$, and a dinner from set $D$ where:

$$\begin{aligned} B &= \{\text{pancakes, bacon and eggs, bagel, Doritos}\} \\ L &= \{\text{burger and fries, garden salad, Doritos}\} \\ D &= \{\text{macaroni, pizza, frozen burrito, pasta, Doritos}\} \end{aligned}$$

Then $B \times L \times D$ is the set of all possible daily diets. Here are some sample elements:

(pancakes, burger and fries, pizza)

(bacon and eggs, garden salad, pasta)

(Doritos, Doritos, frozen burrito)

The Product Rule tells us how many different daily diets are possible:

$$\begin{aligned} |B \times L \times D| &= |B| \cdot |L| \cdot |D| \\ &= 4 \cdot 3 \cdot 5 \\ &= 60. \end{aligned}$$

### 15.2.2 Subsets of an $n$-element Set

The fact that there are $2^n$ subsets of an $n$-element set was proved in Theorem 5.1.5 by setting up a bijection between the subsets and the length-$n$ bit-strings. So the original problem about subsets was tranformed into a question about sequences — *exactly according to plan!*. Now we can fill in the missing explanation of why there are $2^n$ length-$n$ bit-strings: we can write the set of all $n$-bit sequences as a product of sets:

$$\{0,1\}^n ::= \underbrace{\{0,1\} \times \{0,1\} \times \ldots \times \{0,1\}}_{n \text{ terms}}.$$

Then Product Rule gives the answer:

$$|\{0,1\}^n| = |\{0,1\}|^n = 2^n.$$

### 15.2.3 The Sum Rule

Bart allocates his little sister Lisa a quota of 20 crabby days, 40 irritable days, and 60 generally surly days. On how many days can Lisa be out-of-sorts one way or another? Let set $C$ be her crabby days, $I$ be her irritable days, and $S$ be the generally surly. In these terms, the answer to the question is $|C \cup I \cup S|$. Now assuming that she is permitted at most one bad quality each day, the size of this union of sets is given by the *Sum Rule*:

**Rule 15.2.2** (Sum Rule)**.** *If $A_1, A_2, \ldots, A_n$ are* disjoint *sets, then:*

$$|A_1 \cup A_2 \cup \ldots \cup A_n| = |A_1| + |A_2| + \ldots + |A_n|$$

Thus, according to Bart's budget, Lisa can be out-of-sorts for:

$$\begin{aligned} |C \cup I \cup S| &= |C| + |I| + |S| \\ &= 20 + 40 + 60 \\ &= 120 \text{ days} \end{aligned}$$

Notice that the Sum Rule holds only for a union of *disjoint* sets. Finding the size of a union of overlapping sets is a more complicated problem that we'll take up in Section 15.12.

### 15.2.4 Counting Passwords

Few counting problems can be solved with a single rule. More often, a solution is a flurry of sums, products, bijections, and other methods.

For solving problems involving passwords, telephone numbers, and license plates, the sum and product rules are useful together. For example, on a certain computer system, a valid password is a sequence of between six and eight symbols. The first symbol must be a letter (which can be lowercase or uppercase), and the remaining symbols must be either letters or digits. How many different passwords are possible?

Let's define two sets, corresponding to valid symbols in the first and subsequent positions in the password.

$$\begin{aligned} F &= \{a, b, \ldots, z, A, B, \ldots, Z\} \\ S &= \{a, b, \ldots, z, A, B, \ldots, Z, 0, 1, \ldots, 9\} \end{aligned}$$

In these terms, the set of all possible passwords is:[1]

$$(F \times S^5) \cup (F \times S^6) \cup (F \times S^7)$$

[1] The notation $S^5$ means $S \times S \times S \times S \times S$.

Thus, the length-six passwords are in the set $F \times S^5$, the length-seven passwords are in $F \times S^6$, and the length-eight passwords are in $F \times S^7$. Since these sets are disjoint, we can apply the Sum Rule and count the total number of possible passwords as follows:

$$\begin{aligned}
&|(F \times S^5) \cup (F \times S^6) \cup (F \times S^7)| \\
&\quad = |F \times S^5| + |F \times S^6| + |F \times S^7| && \text{Sum Rule} \\
&\quad = |F| \cdot |S|^5 + |F| \cdot |S|^6 + |F| \cdot |S|^7 && \text{Product Rule} \\
&\quad = 52 \cdot 62^5 + 52 \cdot 62^6 + 52 \cdot 62^7 \\
&\quad \approx 1.8 \cdot 10^{14} \text{ different passwords.}
\end{aligned}$$

## 15.3 The Generalized Product Rule

In how many ways can, say, a Nobel prize, a Japan prize, and a Pulitzer prize be awarded to $n$ people? This is easy to answer using our strategy of translating the problem about awards into a problem about sequences. Let $P$ be the set of $n$ people taking the course. Then there is a bijection from ways of awarding the three prizes to the set $P^3 ::= P \times P \times P$. In particular, the assignment:

"Barak wins a Nobel, George wins a Japan, and Bill wins a Pulitzer prize"

maps to the sequence (Barak, George, Bill). By the Product Rule, we have $|P^3| = |P|^3 = n^3$, so there are $n^3$ ways to award the prizes to a class of $n$ people. Notice that $P^3$ includes triples like (Barak, Bill, Barak) where one person wins more than one prize.

But what if the three prizes must be awarded to *different* students? As before, we could map first assignment to the triple (Bill, George, Barak) $\in P^3$. But this function is *no longer a bijection*. For example, no valid assignment maps to the triple (Barak, Bill, Barak) because now we're not allowing Barak to receive two prize. However, there *is* a bijection from prize assignments to the set:

$$S = \{(x, y, z) \in P^3 \mid x, y, \text{and } z \text{ are different people}\}$$

This reduces the original problem to a problem of counting sequences. Unfortunately, the Product Rule does not apply directly to counting sequences of this type because the entries depend on one another; in particular, they must all be different. However, a slightly sharper tool does the trick.

## Prizes for *truly exceptional* Coursework

Given everyone's hard work on this material, the instructors considered awarding some prizes for truly exceptional coursework. Here are three possible prize categories:

**Best Administrative Critique** We asserted that the quiz was closed-book. On the cover page, one strong candidate for this award wrote, "There is no book."

**Awkward Question Award** "Okay, the left sock, right sock, and pants are in an antichain, but how —even with assistance —could I put on all three at once?"

**Best Collaboration Statement** Inspired by a student who wrote "I worked alone" on Quiz 1.

**Rule 15.3.1** (Generalized Product Rule)**.** *Let $S$ be a set of length-$k$ sequences. If there are:*

- *$n_1$ possible first entries,*
- *$n_2$ possible second entries for each first entry,*
  $\vdots$
- *$n_k$ possible $k$th entries for each sequence of first $k-1$ entries,*

*then:*

$$|S| = n_1 \cdot n_2 \cdot n_3 \cdots n_k$$

In the awards example, $S$ consists of sequences $(x, y, z)$. There are $n$ ways to choose $x$, the recipient of prize #1. For each of these, there are $n-1$ ways to choose $y$, the recipient of prize #2, since everyone except for person $x$ is eligible. For each combination of $x$ and $y$, there are $n-2$ ways to choose $z$, the recipient of prize #3, because everyone except for $x$ and $y$ is eligible. Thus, according to the Generalized Product Rule, there are

$$|S| = n \cdot (n-1) \cdot (n-2)$$

ways to award the 3 prizes to different people.

### 15.3.1 Defective Dollar Bills

A dollar bill is *defective* if some digit appears more than once in the 8-digit serial number. If you check your wallet, you'll be sad to discover that defective bills are all-too-common. In fact, how common are *nondefective* bills? Assuming that the digit portions of serial numbers all occur equally often, we could answer this question by computing

$$\text{fraction of nondefective bills} = \frac{|\{\text{serial \#'s with all digits different}\}|}{|\{\text{serial numbers}\}|}. \tag{15.1}$$

Let's first consider the denominator. Here there are no restrictions; there are 10 possible first digits, 10 possible second digits, 10 third digits, and so on. Thus, the total number of 8-digit serial numbers is $10^8$ by the Product Rule.

Next, let's turn to the numerator. Now we're not permitted to use any digit twice. So there are still 10 possible first digits, but only 9 possible second digits, 8 possible third digits, and so forth. Thus, by the Generalized Product Rule, there are

$$10 \cdot 9 \cdot 8 \cdot 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 = \frac{10!}{2} = 1{,}814{,}400$$

serial numbers with all digits different. Plugging these results into Equation 15.1, we find:

$$\text{fraction of nondefective bills} = \frac{1{,}814{,}400}{100{,}000{,}000} = 1.8144\%$$

### 15.3.2 A Chess Problem

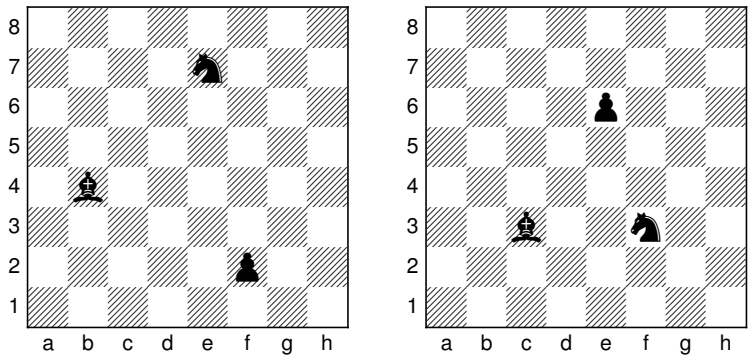
In how many different ways can we place a pawn ($P$), a knight ($N$), and a bishop ($B$) on a chessboard so that no two pieces share a row or a column? A valid configuration is shown in Figure 15.1(a), and an invalid configuration is shown in Figure 15.1(b).

First, we map this problem about chess pieces to a question about sequences. There is a bijection from configurations to sequences

$$(r_P, c_P, r_N, c_N, r_B, c_B)$$

where $r_P$, $r_N$, and $r_B$ are distinct rows and $c_P$, $c_N$, and $c_B$ are distinct columns. In particular, $r_P$ is the pawn's row, $c_P$ is the pawn's column, $r_N$ is the knight's row, etc. Now we can count the number of such sequences using the Generalized Product Rule:

- $r_P$ is one of 8 rows




(a) valid (b) invalid

**Figure 15.1** Two ways of placing a pawn (♙), a knight (♘), and a bishop (♗) on a chessboard. The configuration shown in (b) is invalid because the bishop and the knight are in the same row.

- $c_P$ is one of 8 columns
- $r_N$ is one of 7 rows (any one but $r_P$)
- $c_N$ is one of 7 columns (any one but $c_P$)
- $r_B$ is one of 6 rows (any one but $r_P$ or $r_N$)
- $c_B$ is one of 6 columns (any one but $c_P$ or $c_N$)

Thus, the total number of configurations is $(8 \cdot 7 \cdot 6)^2$.

### 15.3.3 Permutations

A *permutation* of a set $S$ is a sequence that contains every element of $S$ exactly once. For example, here are all the permutations of the set $\{a, b, c\}$:

$$\begin{array}{lll} (a,b,c) & (a,c,b) & (b,a,c) \\ (b,c,a) & (c,a,b) & (c,b,a) \end{array}$$

How many permutations of an $n$-element set are there? Well, there are $n$ choices for the first element. For each of these, there are $n-1$ remaining choices for the second element. For every combination of the first two elements, there are $n-2$ ways to choose the third element, and so forth. Thus, there are a total of

$$n \cdot (n-1) \cdot (n-2) \cdots 3 \cdot 2 \cdot 1 = n!$$

permutations of an $n$-element set. In particular, this formula says that there are

$3! = 6$ permutations of the 3-element set $\{a, b, c\}$, which is the number we found above.

Permutations will come up again in this course approximately 1.6 bazillion times. In fact, permutations are the reason why factorial comes up so often and why we taught you Stirling's approximation:

$$n! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n.$$

## 15.4 The Division Rule

Counting ears and dividing by two is a silly way to count the number of people in a room, but this approach is representative of a powerful counting principle.

A *k-to-1 function* maps exactly $k$ elements of the domain to every element of the codomain. For example, the function mapping each ear to its owner is 2-to-1. Similarly, the function mapping each finger to its owner is 10-to-1, and the function mapping each finger and toe to its owner is 20-to-1. The general rule is:

**Rule 15.4.1** (Division Rule)**.** *If $f : A \to B$ is $k$-to-1, then $|A| = k \cdot |B|$.*

For example, suppose $A$ is the set of ears in the room and $B$ is the set of people. There is a 2-to-1 mapping from ears to people, so by the Division Rule, $|A| = 2 \cdot |B|$. Equivalently, $|B| = |A|/2$, expressing what we knew all along: the number of people is half the number of ears. Unlikely as it may seem, many counting problems are made much easier by initially counting every item multiple times and then correcting the answer using the Division Rule. Let's look at some examples.

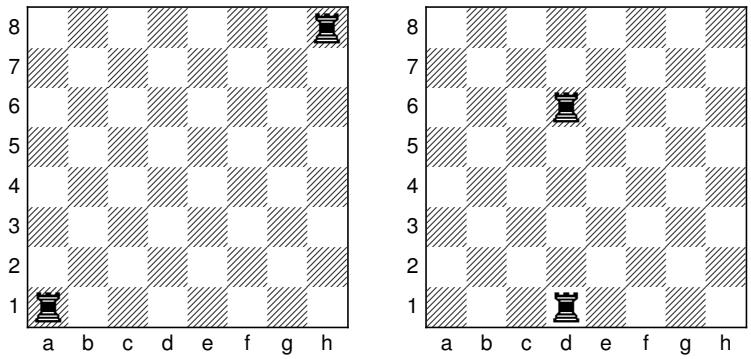
### 15.4.1 Another Chess Problem

In how many different ways can you place two identical rooks on a chessboard so that they do not share a row or column? A valid configuration is shown in Figure 15.2(a), and an invalid configuration is shown in Figure 15.2(b).

Let $A$ be the set of all sequences

$$(r_1, c_1, r_2, c_2)$$

where $r_1$ and $r_2$ are distinct rows and $c_1$ and $c_2$ are distinct columns. Let $B$ be the set of all valid rook configurations. There is a natural function $f$ from set $A$ to set $B$; in particular, $f$ maps the sequence $(r_1, c_1, r_2, c_2)$ to a configuration with one rook in row $r_1$, column $c_1$ and the other rook in row $r_2$, column $c_2$.




(a) valid (b) invalid

**Figure 15.2** Two ways to place 2 rooks (♖) on a chessboard. The configuration in (b) is invalid because the rooks are in the same column.

But now there's a snag. Consider the sequences:

$$(1, 1, 8, 8) \qquad \text{and} \qquad (8, 8, 1, 1)$$

The first sequence maps to a configuration with a rook in the lower-left corner and a rook in the upper-right corner. The second sequence maps to a configuration with a rook in the upper-right corner and a rook in the lower-left corner. The problem is that those are two different ways of describing the *same* configuration! In fact, this arrangement is shown in Figure 15.2(a).

More generally, the function $f$ maps exactly two sequences to *every* board configuration; that is $f$ is a 2-to-1 function. Thus, by the quotient rule, $|A| = 2 \cdot |B|$. Rearranging terms gives:

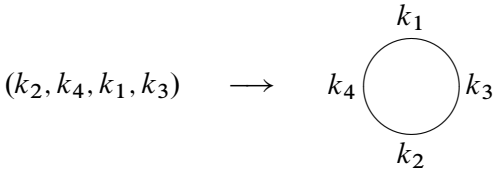
$$|B| = \frac{|A|}{2} = \frac{(8 \cdot 7)^2}{2}.$$

On the second line, we've computed the size of $A$ using the General Product Rule just as in the earlier chess problem.

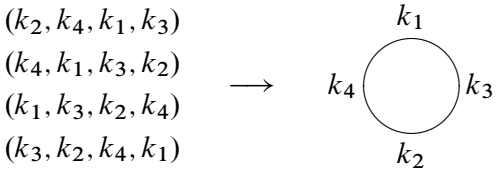
### 15.4.2 Knights of the Round Table

In how many ways can King Arthur arrange to seat his $n$ different knights at his round table? Two seatings are considered to be the same *arrangement* if they yield the same sequence of knights starting at knight number 1 and going clockwise around the table. For example, the following two seatings determine the same arrangement:


So a seating is determined by the sequence of knights going clockwise around the table starting at the top seat. This means seatings are formally the same as the set, $A$, of all permutations of the knights. An arrangement is determined by the sequence of knights going clockwise around the table starting after knight number 1, so it is formally the same as the set, $B$, of all permutations of knights 2 through $n$. We can map each permutation in $A$ to an arrangement in set $B$ by seating the first knight in the permutation at the top of the table, putting the second knight to his left, the third knight to the left of the second, and so forth all the way around the table. For example:




This mapping is actually an $n$-to-1 function from $A$ to $B$, since all $n$ cyclic shifts of the original sequence map to the same seating arrangement. In the example, $n = 4$ different sequences map to the same seating arrangement:




Therefore, by the division rule, the number of circular seating arrangements is:

$$|B| = \frac{|A|}{n} = \frac{n!}{n} = (n-1)!$$

Note that $|A| = n!$ since there are $n!$ permutations of $n$ knights.

## 15.5 Counting Subsets

How many $k$-element subsets of an $n$-element set are there? This question arises all the time in various guises:

- In how many ways can I select 5 books from my collection of 100 to bring on vacation?
- How many different 13-card Bridge hands can be dealt from a 52-card deck?
- In how many ways can I select 5 toppings for my pizza if there are 14 available toppings?

This number comes up so often that there is a special notation for it:

$$\binom{n}{k} ::= \text{ the number of } k\text{-element subsets of an } n\text{-element set.}$$

The expression $\binom{n}{k}$ is read "$n$ choose $k$." Now we can immediately express the answers to all three questions above:

- I can select 5 books from 100 in $\binom{100}{5}$ ways.
- There are $\binom{52}{13}$ different Bridge hands.
- There are $\binom{14}{5}$ different 5-topping pizzas, if 14 toppings are available.

### 15.5.1 The Subset Rule

We can derive a simple formula for the $n$ choose $k$ number using the Division Rule. We do this by mapping any permutation of an $n$-element set $\{a_1, \ldots, a_n\}$ into a $k$-element subset simply by taking the first $k$ elements of the permutation. That is, the permutation $a_1 a_2 \ldots a_n$ will map to the set $\{a_1, a_2, \ldots, a_k\}$.

Notice that any other permutation with the same first $k$ elements $a_1, \ldots, a_k$ *in any order* and the same remaining elements $n - k$ elements *in any order* will also map to this set. What's more, a permutation can only map to $\{a_1, a_2, \ldots, a_k\}$ if its first $k$ elements are the elements $a_1, \ldots, a_k$ in some order. Since there are $k!$ possible permutations of the first $k$ elements and $(n-k)!$ permutations of the remaining elements, we conclude from the Product Rule that exactly $k!(n-k)!$ permutations of the $n$-element set map to the the particular subset, $S$. In other words, the mapping from permutations to $k$-element subsets is $k!(n-k)!$-to-1.

But we know there are $n!$ permutations of an $n$-element set, so by the Division Rule, we conclude that

$$n! = k!(n-k)!\binom{n}{k}$$

which proves:

**Rule 15.5.1** (Subset Rule)**.** *The number of $k$-element subsets of an $n$-element set is*

$$\binom{n}{k} = \frac{n!}{k!\,(n-k)!}.$$

Notice that this works even for 0-element subsets: $n!/0!n! = 1$. Here we use the fact that 0! is a *product* of 0 terms, which by convention[2] equals 1.

### 15.5.2 Bit Sequences

How many $n$-bit sequences contain exactly $k$ ones? We've already seen the straightforward bijection between subsets of an $n$-element set and $n$-bit sequences. For example, here is a 3-element subset of $\{x_1, x_2, \ldots, x_8\}$ and the associated 8-bit sequence:

$$\begin{array}{llllllllll} \{ & x_1, & & & x_4, & x_5 & & & & \} \\ ( & 1, & 0, & 0, & 1, & 1, & 0, & 0, & 0 & ) \end{array}$$

Notice that this sequence has exactly 3 ones, each corresponding to an element of the 3-element subset. More generally, the $n$-bit sequences corresponding to a $k$-element subset will have exactly $k$ ones. So by the Bijection Rule,

**Corollary.** *The number of $n$-bit sequences with exactly $k$ ones is* $\binom{n}{k}$.

## 15.6 Sequences with Repetitions

### 15.6.1 Sequences of Subsets

Choosing a $k$-element subset of an $n$-element set is the same as splitting the set into a pair of subsets: the first subset of size $k$ and the second subset consisting of the remaining $n-k$ elements. So the Subset Rule can be understood as a rule for counting the number of such splits into pairs of subsets.

[2] We don't use it here, but a *sum* of zero terms equals 0.

We can generalize this to splits into more than two subsets. Namely, let $A$ be an $n$-element set and $k_1, k_2, \ldots, k_m$ be nonnegative integers whose sum is $n$. A $(k_1, k_2, \ldots, k_m)$-*split of* $A$ is a sequence

$$(A_1, A_2, \ldots, A_m)$$

where the $A_i$ are disjoint subsets of $A$ and $|A_i| = k_i$ for $i = 1, \ldots, m$.

To count the number of splits we take the same approach as for the Subset Rule. Namely, we map any permutation $a_1 a_2 \ldots a_n$ of an $n$-element set $A$ into a $(k_1, k_2, \ldots, k_m)$-split by letting the 1st subset in the split be the first $k_1$ elements of the permutation, the 2nd subset of the split be the next $k_2$ elements, ..., and the $m$th subset of the split be the final $k_m$ elements of the permutation. This map is a $k_1!\, k_2! \cdots k_m!$-to-1 function from the $n!$ permutations to the $(k_1, k_2, \ldots, k_m)$-splits of $A$, so from the Division Rule we conclude the Subset Split Rule:

**Definition 15.6.1.** For $n, k_1, \ldots, k_m \in \mathbb{N}$, such that $k_1 + k_2 + \cdots + k_m = n$, define the *multinomial coefficient*

$$\binom{n}{k_1, k_2, \ldots, k_m} ::= \frac{n!}{k_1!\, k_2! \ldots k_m!}.$$

**Rule 15.6.2** (Subset Split Rule)**.** *The number of* $(k_1, k_2, \ldots, k_m)$*-splits of an* $n$*-element set is*

$$\binom{n}{k_1, \ldots, k_m}.$$

### 15.6.2 The Bookkeeper Rule

We can also generalize our count of $n$-bit sequences with $k$ ones to counting sequences of $n$ letters over an alphabet with more than two letters. For example, how many sequences can be formed by permuting the letters in the 10-letter word BOOKKEEPER?

Notice that there are 1 B, 2 O's, 2 K's, 3 E's, 1 P, and 1 R in BOOKKEEPER. This leads to a straightforward bijection between permutations of BOOKKEEPER and (1,2,2,3,1,1)-splits of $\{1, 2, \ldots, 10\}$. Namely, map a permutation to the sequence of sets of positions where each of the different letters occur.

For example, in the permutation BOOKKEEPER itself, the B is in the 1st position, the O's occur in the 2nd and 3rd positions, K's in 4th and 5th, the E's in the 6th, 7th and 9th, P in the 8th, and R is in the 10th position. So BOOKKEEPER maps to

$$(\{1\}, \{2, 3\}, \{4, 5\}, \{6, 7, 9\}, \{8\}, \{10\}).$$

From this bijection and the Subset Split Rule, we conclude that the number of ways to rearrange the letters in the word BOOKKEEPER is:

$$\frac{\overbrace{10!}^{\text{total letters}}}{\underbrace{1!}_{\text{B's}}\ \underbrace{2!}_{\text{O's}}\ \underbrace{2!}_{\text{K's}}\ \underbrace{3!}_{\text{E's}}\ \underbrace{1!}_{\text{P's}}\ \underbrace{1!}_{\text{R's}}}$$

This example generalizes directly to an exceptionally useful counting principle which we will call the

**Rule 15.6.3** (Bookkeeper Rule)**.** *Let $l_1, \ldots, l_m$ be distinct elements. The number of sequences with $k_1$ occurrences of $l_1$, and $k_2$ occurrences of $l_2$, ..., and $k_m$ occurrences of $l_m$ is*

$$\binom{k_1 + k_2 + \cdots + k_m}{k_1, \ldots, k_m}.$$

For example, suppose you are planning a 20-mile walk, which should include 5 northward miles, 5 eastward miles, 5 southward miles, and 5 westward miles. How many different walks are possible?

There is a bijection between such walks and sequences with 5 N's, 5 E's, 5 S's, and 5 W's. By the Bookkeeper Rule, the number of such sequences is:

$$\frac{20!}{(5!)^4}.$$

## 15.7 The Binomial Theorem

Counting gives insight into one of the basic theorems of algebra. A *binomial* is a sum of two terms, such as $a + b$. Now consider its 4th power, $(a + b)^4$.

If we multiply out this 4th power expression completely, we get

$$\begin{aligned}(a+b)^4 = \quad & aaaa + aaab + aaba + aabb \\ & + abaa + abab + abba + abbb \\ & + baaa + baab + baba + babb \\ & + bbaa + bbab + bbba + bbbb\end{aligned}$$

Notice that there is one term for every sequence of $a$'s and $b$'s. So there are $2^4$ terms, and the number of terms with $k$ copies of $b$ and $n - k$ copies of $a$ is:

$$\frac{n!}{k!\,(n-k)!} = \binom{n}{k}$$

by the Bookkeeper Rule. Hence, the coefficient of $a^{n-k}b^k$ is $\binom{n}{k}$. So for $n = 4$, this means:

$$(a+b)^4 = \binom{4}{0}\cdot a^4b^0 + \binom{4}{1}\cdot a^3b^1 + \binom{4}{2}\cdot a^2b^2 + \binom{4}{3}\cdot a^1b^3 + \binom{4}{4}\cdot a^0b^4$$

In general, this reasoning gives the Binomial Theorem:

**Theorem 15.7.1** (Binomial Theorem)**.** *For all $n \in \mathbb{N}$ and $a, b \in \mathbb{R}$:*

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^{n-k}b^k$$

The Binomial Theorem explains why the $n$ choose $k$ number is called a *binomial coefficient*.

This reasoning about binomials extends nicely to *multinomials*, which are sums of two or more terms. For example, suppose we wanted the coefficient of

$$bo^2k^2e^3pr$$

in the expansion of $(b+o+k+e+p+r)^{10}$. Each term in this expansion is a product of 10 variables where each variable is one of $b$, $o$, $k$, $e$, $p$, or $r$. Now, the coefficient of $bo^2k^2e^3pr$ is the number of those terms with exactly 1 $b$, 2 $o$'s, 2 $k$'s, 3 $e$'s, 1 $p$, and 1 $r$. And the number of such terms is precisely the number of rearrangements of the word BOOKKEEPER:

$$\binom{10}{1,2,2,3,1,1} = \frac{10!}{1!\,2!\,2!\,3!\,1!\,1!}.$$

This reasoning extends to a general theorem.

**Theorem 15.7.2** (Multinomial Theorem)**.** *For all $n \in \mathbb{N}$,*

$$(z_1 + z_2 + \cdots + z_m)^n = \sum_{\substack{k_1,\ldots,k_m \in \mathbb{N} \\ k_1+\cdots+k_m=n}} \binom{n}{k_1, k_2, \ldots, k_m} z_1^{k_1} z_2^{k_2} \cdots z_m^{k_m}.$$

You'll be better off remembering the reasoning behind the Multinomial Theorem rather than this cumbersome formal statement.

## 15.8 A Word about Words

Someday you might refer to the Subset Split Rule or the Bookkeeper Rule in front of a roomful of colleagues and discover that they're all staring back at you blankly. This is not because they're dumb, but rather because we made up the name "Bookkeeper Rule." However, the rule is excellent and the name is apt, so we suggest that you play through: "You know? The Bookkeeper Rule? Don't you guys know *anything???*"

The Bookkeeper Rule is sometimes called the "formula for permutations with indistinguishable objects." The size $k$ subsets of an $n$-element set are sometimes called *$k$-combinations*. Other similar-sounding descriptions are "combinations with repetition, permutations with repetition, $r$-permutations, permutations with indistinguishable objects," and so on. However, the counting rules we've taught you are sufficient to solve all these sorts of problems without knowing this jargon, so we won't burden you with it.

## 15.9 Counting Practice: Poker Hands

Five-Card Draw is a card game in which each player is initially dealt a *hand* consisting of 5 cards from a deck of 52 cards.[3] (Then the game gets complicated, but let's not worry about that.) The number of different hands in Five-Card Draw is the number of 5-element subsets of a 52-element set, which is

$$\binom{52}{5} = 2,598,960.$$

Let's get some counting practice by working out the number of hands with various special properties.

[3] There are 52 cards in a standard deck. Each card has a *suit* and a *rank*. There are four suits:

♠ (spades) ♡ (hearts) ♣ (clubs) ◇ (diamonds)

And there are 13 ranks, listed here from lowest to highest:

$$\overset{\text{Ace}}{A}, 2, 3, 4, 5, 6, 7, 8, 9, \overset{\text{Jack}}{J}, \overset{\text{Queen}}{Q}, \overset{\text{King}}{K}.$$

Thus, for example, 8♡ is the 8 of hearts and $A$♠ is the ace of spades.

### 15.9.1 Hands with a Four-of-a-Kind

A *Four-of-a-Kind* is a set of four cards with the same rank. How many different hands contain a Four-of-a-Kind? Here are a couple examples:

$$\{8\spadesuit,\ 8\diamondsuit,\ Q\heartsuit,\ 8\heartsuit,\ 8\clubsuit\}$$
$$\{A\clubsuit,\ 2\clubsuit,\ 2\heartsuit,\ 2\diamondsuit,\ 2\spadesuit\}$$

As usual, the first step is to map this question to a sequence-counting problem. A hand with a Four-of-a-Kind is completely described by a sequence specifying:

1. The rank of the four cards.
2. The rank of the extra card.
3. The suit of the extra card.

Thus, there is a bijection between hands with a Four-of-a-Kind and sequences consisting of two distinct ranks followed by a suit. For example, the three hands above are associated with the following sequences:

$$(8, Q, \heartsuit) \leftrightarrow \{8\spadesuit,\ 8\diamondsuit,\ 8\heartsuit,\ 8\clubsuit,\ Q\heartsuit\}$$
$$(2, A, \clubsuit) \leftrightarrow \{2\clubsuit,\ 2\heartsuit,\ 2\diamondsuit,\ 2\spadesuit,\ A\clubsuit\}$$

Now we need only count the sequences. There are 13 ways to choose the first rank, 12 ways to choose the second rank, and 4 ways to choose the suit. Thus, by the Generalized Product Rule, there are $13 \cdot 12 \cdot 4 = 624$ hands with a Four-of-a-Kind. This means that only 1 hand in about 4165 has a Four-of-a-Kind. Not surprisingly, Four-of-a-Kind is considered to be a very good poker hand!

### 15.9.2 Hands with a Full House

A *Full House* is a hand with three cards of one rank and two cards of another rank. Here are some examples:

$$\{2\spadesuit,\ 2\clubsuit,\ 2\diamondsuit,\ J\clubsuit,\ J\diamondsuit\}$$
$$\{5\diamondsuit,\ 5\clubsuit,\ 5\heartsuit,\ 7\heartsuit,\ 7\clubsuit\}$$

Again, we shift to a problem about sequences. There is a bijection between Full Houses and sequences specifying:

1. The rank of the triple, which can be chosen in 13 ways.
2. The suits of the triple, which can be selected in $\binom{4}{3}$ ways.
3. The rank of the pair, which can be chosen in 12 ways.
4. The suits of the pair, which can be selected in $\binom{4}{2}$ ways.

The example hands correspond to sequences as shown below:

$$(2, \{\spadesuit, \clubsuit, \diamondsuit\}, J, \{\clubsuit, \diamondsuit\}) \leftrightarrow \{2\spadesuit,\ 2\clubsuit,\ 2\diamondsuit,\ J\clubsuit,\ J\diamondsuit\}$$
$$(5, \{\diamondsuit, \clubsuit, \heartsuit\}, 7, \{\heartsuit, \clubsuit\}) \leftrightarrow \{5\diamondsuit,\ 5\clubsuit,\ 5\heartsuit,\ 7\heartsuit,\ 7\clubsuit\}$$

By the Generalized Product Rule, the number of Full Houses is:

$$13 \cdot \binom{4}{3} \cdot 12 \cdot \binom{4}{2}.$$

We're on a roll —but we're about to hit a speed bump.

### 15.9.3 Hands with Two Pairs

How many hands have *Two Pairs*; that is, two cards of one rank, two cards of another rank, and one card of a third rank? Here are examples:

$$\{3\diamondsuit,\ 3\spadesuit,\ Q\diamondsuit,\ Q\heartsuit,\ A\clubsuit\}$$
$$\{9\heartsuit,\ 9\diamondsuit,\ 5\heartsuit,\ 5\clubsuit,\ K\spadesuit\}$$

Each hand with Two Pairs is described by a sequence consisting of:

1. The rank of the first pair, which can be chosen in 13 ways.
2. The suits of the first pair, which can be selected $\binom{4}{2}$ ways.
3. The rank of the second pair, which can be chosen in 12 ways.
4. The suits of the second pair, which can be selected in $\binom{4}{2}$ ways.
5. The rank of the extra card, which can be chosen in 11 ways.
6. The suit of the extra card, which can be selected in $\binom{4}{1} = 4$ ways.

Thus, it might appear that the number of hands with Two Pairs is:

$$13 \cdot \binom{4}{2} \cdot 12 \cdot \binom{4}{2} \cdot 11 \cdot 4.$$

Wrong answer! The problem is that there is *not* a bijection from such sequences to hands with Two Pairs. This is actually a 2-to-1 mapping. For example, here are the pairs of sequences that map to the hands given above:

$$\begin{array}{ll} (3, \{\diamondsuit, \spadesuit\}, Q, \{\diamondsuit, \heartsuit\}, A, \clubsuit) & \searrow \\ & \{3\diamondsuit,\ 3\spadesuit,\ Q\diamondsuit,\ Q\heartsuit,\ A\clubsuit\} \\ (Q, \{\diamondsuit, \heartsuit\}, 3, \{\diamondsuit, \spadesuit\}, A, \clubsuit) & \nearrow \\ & \\ (9, \{\heartsuit, \diamondsuit\}, 5, \{\heartsuit, \clubsuit\}, K, \spadesuit) & \searrow \\ & \{9\heartsuit,\ 9\diamondsuit,\ 5\heartsuit,\ 5\clubsuit,\ K\spadesuit\} \\ (5, \{\heartsuit, \clubsuit\}, 9, \{\heartsuit, \diamondsuit\}, K, \spadesuit) & \nearrow \end{array}$$

The problem is that nothing distinguishes the first pair from the second. A pair of 5's and a pair of 9's is the same as a pair of 9's and a pair of 5's. We avoided this difficulty in counting Full Houses because, for example, a pair of 6's and a triple of kings is different from a pair of kings and a triple of 6's.

We ran into precisely this difficulty last time, when we went from counting arrangements of *different* pieces on a chessboard to counting arrangements of two *identical* rooks. The solution then was to apply the Division Rule, and we can do the same here. In this case, the Division rule says there are twice as many sequences as hands, so the number of hands with Two Pairs is actually:

$$\frac{13 \cdot \binom{4}{2} \cdot 12 \cdot \binom{4}{2} \cdot 11 \cdot 4}{2}.$$

### Another Approach

The preceding example was disturbing! One could easily overlook the fact that the mapping was 2-to-1 on an exam, fail the course, and turn to a life of crime. You can make the world a safer place in two ways:

1. Whenever you use a mapping $f : A \to B$ to translate one counting problem to another, check that the same number elements in $A$ are mapped to each element in $B$. If $k$ elements of $A$ map to each of element of $B$, then apply the Division Rule using the constant $k$.

2. As an extra check, try solving the same problem in a different way. Multiple approaches are often available —and all had better give the same answer!

(Sometimes different approaches give answers that *look* different, but turn out to be the same after some algebra.)

We already used the first method; let's try the second. There is a bijection between hands with two pairs and sequences that specify:

1. The ranks of the two pairs, which can be chosen in $\binom{13}{2}$ ways.
2. The suits of the lower-rank pair, which can be selected in $\binom{4}{2}$ ways.
3. The suits of the higher-rank pair, which can be selected in $\binom{4}{2}$ ways.
4. The rank of the extra card, which can be chosen in 11 ways.
5. The suit of the extra card, which can be selected in $\binom{4}{1} = 4$ ways.

For example, the following sequences and hands correspond:

$$(\{3, Q\}, \{\diamondsuit, \spadesuit\}, \{\diamondsuit, \heartsuit\}, A, \clubsuit) \leftrightarrow \{3\diamondsuit,\ 3\spadesuit,\ Q\diamondsuit,\ Q\heartsuit,\ A\clubsuit\}$$
$$(\{9, 5\}, \{\heartsuit, \clubsuit\}, \{\heartsuit, \diamondsuit\}, K, \spadesuit) \leftrightarrow \{9\heartsuit,\ 9\diamondsuit,\ 5\heartsuit,\ 5\clubsuit,\ K\spadesuit\}$$

Thus, the number of hands with two pairs is:

$$\binom{13}{2} \cdot \binom{4}{2} \cdot \binom{4}{2} \cdot 11 \cdot 4.$$

This is the same answer we got before, though in a slightly different form.

### 15.9.4 Hands with Every Suit

How many hands contain at least one card from every suit? Here is an example of such a hand:

$$\{7\diamondsuit,\ K\clubsuit,\ 3\diamondsuit,\ A\heartsuit,\ 2\spadesuit\}$$

Each such hand is described by a sequence that specifies:

1. The ranks of the diamond, the club, the heart, and the spade, which can be selected in $13 \cdot 13 \cdot 13 \cdot 13 = 13^4$ ways.
2. The suit of the extra card, which can be selected in 4 ways.
3. The rank of the extra card, which can be selected in 12 ways.

For example, the hand above is described by the sequence:

$$(7, K, A, 2, \diamondsuit, 3) \leftrightarrow \{7\diamondsuit,\ K\clubsuit,\ A\heartsuit,\ 2\spadesuit,\ 3\diamondsuit\}.$$

Are there other sequences that correspond to the same hand? There is one more! We could equally well regard either the $3\diamondsuit$ or the $7\diamondsuit$ as the extra card, so this is actually a 2-to-1 mapping. Here are the two sequences corresponding to the example hand:

$$\begin{array}{ll} (7, K, A, 2, \diamondsuit, 3) & \searrow \\ & \{7\diamondsuit,\ K\clubsuit,\ A\heartsuit,\ 2\spadesuit,\ 3\diamondsuit\} \\ (3, K, A, 2, \diamondsuit, 7) & \nearrow \end{array}$$

Therefore, the number of hands with every suit is:

$$\frac{13^4 \cdot 4 \cdot 12}{2}.$$

## 15.10 The Pigeonhole Principle

Here is an old puzzle:

> A drawer in a dark room contains red socks, green socks, and blue socks. How many socks must you withdraw to be sure that you have a matching pair?

For example, picking out three socks is not enough; you might end up with one red, one green, and one blue. The solution relies on the

### Pigeonhole Principle

*If there are more pigeons than holes they occupy, then at least two pigeons must be in the same hole.*

What pigeons have to do with selecting footwear under poor lighting conditions may not be immediately obvious, but if we let socks be pigeons and the colors be three pigeonholes, then as soon as you pick four socks, there are bound to be two in the same hole, that is, with the same color. So four socks are enough to ensure a matched pair. For example, one possible mapping of four socks to three colors is shown in Figure 15.3.

A rigorous statement of the Principle goes this way:

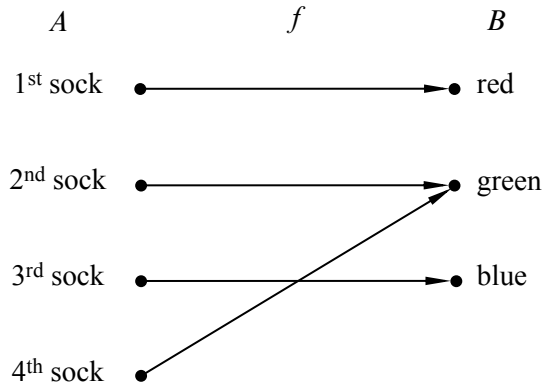



**Figure 15.3** One possible mapping of four socks to three colors.

**Rule 15.10.1** (Pigeonhole Principle)**.** *If $|A| > |B|$, then for every total function $f : A \to B$, there exist two different elements of $A$ that are mapped by $f$ to the same element of $B$.*

Stating the Principle this way may be less intuitive, but it should now sound familiar: it is simply the contrapositive of the Mapping Rules injective case (5.2). Here, the pigeons form set $A$, the pigeonholes are the set $B$, and $f$ describes which hole each pigeon occupies.

Mathematicians have come up with many ingenious applications for the pigeonhole principle. If there were a cookbook procedure for generating such arguments, we'd give it to you. Unfortunately, there isn't one. One helpful tip, though: when you try to solve a problem with the pigeonhole principle, the key is to clearly identify three things:

1. The set $A$ (the pigeons).
2. The set $B$ (the pigeonholes).
3. The function $f$ (the rule for assigning pigeons to pigeonholes).

### 15.10.1 Hairs on Heads

There are a number of generalizations of the pigeonhole principle. For example:

**Rule 15.10.2** (Generalized Pigeonhole Principle)**.** *If $|A| > k \cdot |B|$, then every total function $f : A \to B$ maps at least $k+1$ different elements of $A$ to the same element of $B$.*

For example, if you pick two people at random, surely they are extremely unlikely to have *exactly* the same number of hairs on their heads. However, in the remarkable city of Boston, Massachusetts there are actually *three* people who have exactly the same number of hairs! Of course, there are many bald people in Boston, and they all have zero hairs. But we're talking about non-bald people; say a person is non-bald if they have at least ten thousand hairs on their head.

Boston has about 500,000 non-bald people, and the number of hairs on a person's head is at most 200,000. Let $A$ be the set of non-bald people in Boston, let $B = \{10,000, 10,001, \ldots, 200,000\}$, and let $f$ map a person to the number of hairs on his or her head. Since $|A| > 2|B|$, the Generalized Pigeonhole Principle implies that at least three people have exactly the same number of hairs. We don't know who they are, but we know they exist!

### 15.10.2 Subsets with the Same Sum

For your reading pleasure, we have displayed ninety 25-digit numbers in Figure 15.4. Are there two different subsets of these 25-digit numbers that have the same sum? For example, maybe the sum of the last ten numbers in the first column is equal to the sum of the first eleven numbers in the second column?

Finding two subsets with the same sum may seem like a silly puzzle, but solving these sorts of problems turns out to be useful in diverse applications such as finding good ways to fit packages into shipping containers and decoding secret messages.

It turns out that it is hard to find different subsets with the same sum, which is why this problem arises in cryptography. But it is easy to prove that two such subsets *exist*. That's where the Pigeonhole Principle comes in.

Let $A$ be the collection of all subsets of the 90 numbers in the list. Now the sum of any subset of numbers is at most $90 \cdot 10^{25}$, since there are only 90 numbers and every 25-digit number is less than $10^{25}$. So let $B$ be the set of integers $\{0, 1, \ldots, 90 \cdot 10^{25}\}$, and let $f$ map each subset of numbers (in $A$) to its sum (in $B$).

We proved that an $n$-element set has $2^n$ different subsets in Section 15.2. Therefore:

$$|A| = 2^{90} \geq 1.237 \times 10^{27}$$

On the other hand:

$$|B| = 90 \cdot 10^{25} + 1 \leq 0.901 \times 10^{27}.$$

Both quantities are enormous, but $|A|$ is a bit greater than $|B|$. This means that $f$ maps at least two elements of $A$ to the same element of $B$. In other words, by the Pigeonhole Principle, two different subsets must have the same sum!

Notice that this proof gives no indication *which* two sets of numbers have the same sum. This frustrating variety of argument is called a *nonconstructive proof*.

| | |
|---|---|
| 00204801353855029644480 38 | 31710048321735013941130 17 |
| 57632573310834796474093 98 | 82473310000429953116460 21 |
| 04894459918669156762409 92 | 32082344215973686470192 65 |
| 58009491235489891226286 63 | 84962439971234759227663 10 |
| 10826620324303796513709 81 | 34372546563551578648691 13 |
| 60429008011992802180260 01 | 85183991406760026607474 77 |
| 11784808947697061789949 93 | 35748833903586539237113 65 |
| 61161717891377378967014 05 | 85436912834701914523337 63 |
| 12531273516832396938513 27 | 36449099460404801899691 49 |
| 61448689730015823697235 12 | 86753092583741370924613 52 |
| 13015051292340778110690 11 | 37900441327370840944172 46 |
| 62473145938511692347461 52 | 86943211123639968672966 65 |
| 13115671111438664338821 94 | 38703321274379713553228 15 |
| 68144289442668749634882 74 | 87723212036084772458511 54 |
| 14700294527212035876862 14 | 40805058045778014513631 00 |
| 68708529455438868491478 81 | 87914221617225825463410 91 |
| 15782710472862574994338 86 | 41672834610257023481249 20 |
| 69149555081209500937323 97 | 90626280245921262839732 85 |
| 16382439218521762431923 54 | 42359968311237777882112 49 |
| 69496324513659871524235 41 | 91378455669255263498977 94 |
| 17635802191319859631023 65 | 46709394457494390421112 20 |
| 71282111436136198284156 50 | 91537629668031892919344 19 |
| 18262277956018422310296 94 | 48153793518653842796134 27 |
| 71739200836518623079253 94 | 92708801940776364069842 49 |
| 18439718626751020372014 20 | 48370529482129226044421 90 |
| 72156548742117556762205 87 | 93243014807221034903792 04 |
| 23969511937221345261772 37 | 51063894238550185506715 30 |
| 72569328471643910402330 50 | 94360908321466951471405 81 |
| 27813945682685998010963 54 | 51423681920047692180699 10 |
| 73328226570752354316203 17 | 94753081597345382490132 38 |
| 27966051967136104054080 19 | 51812340961301440840418 56 |
| 74264418295415734449641 39 | 94923766239174869749232 02 |
| 29310163947619752631903 47 | 51982673981256179943913 48 |
| 76321981265318093271863 21 | 95119725587798802882529 79 |
| 29334580582944051551972 96 | 53175929403162312197583 72 |
| 77121544322119128823105 11 | 96024134246191871125522 64 |
| 30755144104909759203153 48 | 53843581267717941283569 47 |
| 78589186642402623566100 10 | 96312171149061292194611 11 |
| 81494367168713711619320 35 | 31576931053251112843219 93 |
| 31114749852527934528600 17 | 54392117122489019954234 41 |
| 78981567867632129631786 79 | 99081898531027533359813 19 |
| 31456215879361201184387 01 | 56103798260928381927604 58 |
| 81475910170375733378486 16 | 99132374763417642998139 87 |
| 31489012556288811031985 49 | 56323175554652286776760 44 |
| 56921683746370196174237 12 | 81760638316825365713067 91 |

**Figure 15.4** Ninety 25-digit numbers. Can you find two different subsets of these numbers that have the same sum?

### The $100 prize for two same-sum subsets

To see if was possible to actually *find* two different subsets of the ninety 25-digit numbers with the same sum, we offered a $100 prize to the first student who did it. We didn't expect to have to pay off this bet, but we underestimated the ingenuity and initiative of the students. One computer science major wrote a program that cleverly searched only among a reasonably small set of "plausible" sets, sorted them by their sums, and actually found a couple with the same sum. He won the prize. A few days later, a math major figured out how to reformulate the sum problem as a "lattice basis reduction" problem; then he found a software package implementing an efficient basis reduction procedure, and using it, he very quickly found lots of pairs of subsets with the same sum. He didn't win the prize, but he got a standing ovation from the class —staff included.

### The $500 Prize for Sets with Distinct Subset Sums

How can we construct a set of $n$ positive integers such that all its subsets have *distinct* sums? One way is to use powers of two:

$$\{1, 2, 4, 8, 16\}$$

This approach is so natural that one suspects all other such sets must involve larger numbers. (For example, we could safely replace 16 by 17, but not by 15.) Remarkably, there are examples involving *smaller* numbers. Here is one:

$$\{6, 9, 11, 12, 13\}$$

One of the top mathematicians of the Twentieth Century, Paul Erdős, conjectured in 1931 that there are no such sets involving *significantly* smaller numbers. More precisely, he conjectured that the largest number in such a set must be greater than $c2^n$ for some constant $c > 0$. He offered $500 to anyone who could prove or disprove his conjecture, but the problem remains unsolved.

## 15.11 A Magic Trick

A Magician sends an Assistant into the audience with a deck of 52 cards while the Magician looks away.

Five audience members each select one card from the deck. The Assistant then gathers up the five cards and holds up four of them so the Magician can see them. The Magician concentrates for a short time and then correctly names the secret, fifth card!

Since we don't really believe the Magician can read minds, we know the Assistant has somehow communicated the secret card to the Magician. Real Magicians and Assistants are not to be trusted, so we expect that the Assistant would secretly signal the Magician with coded phrases or body language, but for this trick they don't have to cheat. In fact, the Magician and Assistant could be kept out of sight of each other while some audience member holds up the 4 cards designated by the Assistant for the Magician to see.

Of course, without cheating, there is still an obvious way the Assistant can communicate to the Magician: he can choose any of the $4! = 24$ permutations of the 4 cards as the order in which to hold up the cards. However, this alone won't quite work: there are 48 cards remaining in the deck, so the Assistant doesn't have enough choices of orders to indicate exactly what the secret card is (though he could narrow it down to two cards).

### 15.11.1 The Secret

The method the Assistant can use to communicate the fifth card exactly is a nice application of what we know about counting and matching.

The Assistant has a second legitimate way to communicate: he can choose *which of the five cards to keep hidden*. Of course, it's not clear how the Magician could determine which of these five possibilities the Assistant selected by looking at the four visible cards, but there is a way, as we'll now explain.

The problem facing the Magician and Assistant is actually a bipartite matching problem. Each vertex on left will correspond to the information available to the Assistant, namely, a *set* of 5 cards. So the set $X$ of left hand vertices will have $\binom{52}{5}$ elements.

Each vertex on right will correspond to the information available to the Magician, namely, a *sequence* of 4 distinct cards. So the set $Y$ of right hand vertices will have $52 \cdot 51 \cdot 50 \cdot 49$ elements. When the audience selects a set of 5 cards, then the Assistant must reveal a sequence of 4 cards from that hand. This constraint is represented by having an edge between a set of 5 cards on the left and a sequence of 4 cards on the right precisely when every card in the sequence is also in the set. This specifies the bipartite graph. Some edges are shown in the diagram in Figure 15.5.

For example,

$$\{8\heartsuit, K\spadesuit, Q\spadesuit, 2\diamondsuit, 6\diamondsuit\} \tag{15.2}$$

is an element of $X$ on the left. If the audience selects this set of 5 cards, then

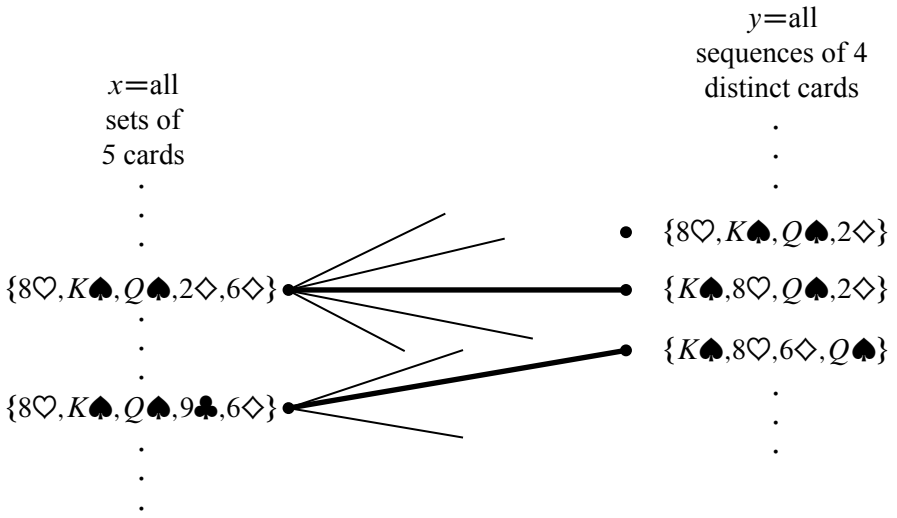



**Figure 15.5** The bipartite graph where the nodes on the left correspond to *sets* of 5 cards and the nodes on the right correspond to *sequences* of 4 cards. There is an edge between a set and a sequence whenever all the cards in the sequence are contained in the set.

there are many different 4-card sequences on the right in set $Y$ that the Assistant could choose to reveal, including $(8\heartsuit, K\spadesuit, Q\spadesuit, 2\diamondsuit)$, $(K\spadesuit, 8\heartsuit, Q\spadesuit, 2\diamondsuit)$, and $(K\spadesuit, 8\heartsuit, 6\diamondsuit, Q\spadesuit)$.

What the Magician and his Assistant need to perform the trick is a *matching* for the $X$ vertices. If they agree in advance on some matching, then when the audience selects a set of 5 cards, the Assistant reveals the matching sequence of 4 cards. The Magician uses the matching to find the audience's chosen set of 5 cards, and so he can name the one not already revealed.

For example, suppose the Assistant and Magician agree on a matching containing the two bold edges in Figure 15.5. If the audience selects the set

$$\{8\heartsuit, K\spadesuit, Q\spadesuit, 9\clubsuit, 6\diamondsuit\}, \tag{15.3}$$

then the Assistant reveals the corresponding sequence

$$(K\spadesuit, 8\heartsuit, 6\diamondsuit, Q\spadesuit). \tag{15.4}$$

Using the matching, the Magician sees that the hand (15.3) is matched to the sequence (15.4), so he can name the one card in the corresponding set not already revealed, namely, the $9\clubsuit$. Notice that the fact that the sets are *matched*, that is, that different sets are paired with *distinct* sequences, is essential. For example, if

the audience picked the previous hand (15.2), it would be possible for the Assistant to reveal the same sequence (15.4), but he better not do that; if he did, then the Magician would have no way to tell if the remaining card was the 9♣ or the 2♢.

So how can we be sure the needed matching can be found? The answer is that each vertex on the left has degree $5 \cdot 4! = 120$, since there are five ways to select the card kept secret and there are 4! permutations of the remaining 4 cards. In addition, each vertex on the right has degree 48, since there are 48 possibilities for the fifth card. So this graph is *degree-constrained* according to Definition 11.5.5, and so has a matching by Theorem 11.5.6.

In fact, this reasoning shows that the Magician could still pull off the trick if 120 cards were left instead of 48, that is, the trick would work with a deck as large as 124 different cards —without any magic!

### 15.11.2 The Real Secret

But wait a minute! It's all very well in principle to have the Magician and his Assistant agree on a matching, but how are they supposed to remember a matching with $\binom{52}{5} = 2,598,960$ edges? For the trick to work in practice, there has to be a way to match hands and card sequences mentally and on the fly.

We'll describe one approach. As a running example, suppose that the audience selects:

$$10\heartsuit \quad 9\diamondsuit \quad 3\heartsuit \quad Q\spadesuit \quad J\diamondsuit.$$

- The Assistant picks out two cards of the same suit. In the example, the assistant might choose the 3♡ and 10♡. This is always possible because of the Pigeonhole Principle —there are five cards and 4 suits so two cards must be in the same suit.

- The Assistant locates the ranks of these two cards on the cycle shown in Figure 15.6. For any two distinct ranks on this cycle, one is always between 1 and 6 hops clockwise from the other. For example, the 3♡ is 6 hops clockwise from the 10♡.

- The more counterclockwise of these two cards is revealed first, and the other becomes the secret card. Thus, in our example, the 10♡ would be revealed, and the 3♡ would be the secret card. Therefore:

  - The suit of the secret card is the same as the suit of the first card revealed.
  - The rank of the secret card is between 1 and 6 hops clockwise from the rank of the first card revealed.

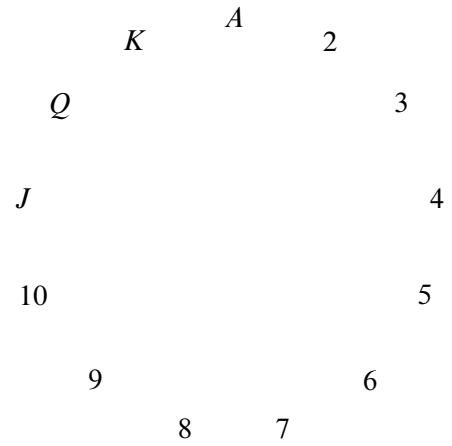



**Figure 15.6** The 13 card ranks arranged in cyclic order.

- All that remains is to communicate a number between 1 and 6. The Magician and Assistant agree beforehand on an ordering of all the cards in the deck from smallest to largest such as:

  $$A\clubsuit\ A\diamondsuit\ A\heartsuit\ A\spadesuit\ 2\clubsuit\ 2\diamondsuit\ 2\heartsuit\ 2\spadesuit\ \ldots\ K\heartsuit\ K\spadesuit$$

  The order in which the last three cards are revealed communicates the number according to the following scheme:

  $$\begin{array}{lll} (\ \text{small}, & \text{medium}, & \text{large}\ ) = 1 \\ (\ \text{small}, & \text{large}, & \text{medium}\ ) = 2 \\ (\ \text{medium}, & \text{small}, & \text{large}\ ) = 3 \\ (\ \text{medium}, & \text{large}, & \text{small}\ ) = 4 \\ (\ \text{large}, & \text{small}, & \text{medium}\ ) = 5 \\ (\ \text{large}, & \text{medium}, & \text{small}\ ) = 6 \end{array}$$

  In the example, the Assistant wants to send 6 and so reveals the remaining three cards in large, medium, small order. Here is the complete sequence that the Magician sees:

  $$10\heartsuit \quad Q\spadesuit \quad J\diamondsuit \quad 9\diamondsuit$$

- The Magician starts with the first card, $10\heartsuit$, and hops 6 ranks clockwise to reach $3\heartsuit$, which is the secret card!

So that's how the trick can work with a standard deck of 52 cards. On the other hand, Hall's Theorem implies that the Magician and Assistant can *in principle* perform the trick with a deck of up to 124 cards. It turns out that there is a method

which they could actually learn to use with a reasonable amount of practice for a 124-card deck, but we won't explain it here.[4]

### 15.11.3 The Same Trick with Four Cards?

Suppose that the audience selects only *four* cards and the Assistant reveals a sequence of *three* to the Magician. Can the Magician determine the fourth card?

Let $X$ be all the sets of four cards that the audience might select, and let $Y$ be all the sequences of three cards that the Assistant might reveal. Now, on one hand, we have

$$|X| = \binom{52}{4} = 270,725$$

by the Subset Rule. On the other hand, we have

$$|Y| = 52 \cdot 51 \cdot 50 = 132,600$$

by the Generalized Product Rule. Thus, by the Pigeonhole Principle, the Assistant must reveal the *same* sequence of three cards for at least

$$\left\lceil \frac{270,725}{132,600} \right\rceil = 3$$

*different* four-card hands. This is bad news for the Magician: if he sees that sequence of three, then there are at least three possibilities for the fourth card which he cannot distinguish. So there is no legitimate way for the Assistant to communicate exactly what the fourth card is!

## 15.12 Inclusion-Exclusion

How big is a union of sets? For example, suppose there are 60 math majors, 200 EECS majors, and 40 physics majors. How many students are there in these three departments? Let $M$ be the set of math majors, $E$ be the set of EECS majors, and $P$ be the set of physics majors. In these terms, we're asking for $|M \cup E \cup P|$.

The Sum Rule says that if $M$, $E$, and $P$ are disjoint, then the sum of their sizes is

$$|M \cup E \cup P| = |M| + |E| + |P|.$$

However, the sets $M$, $E$, and $P$ might *not* be disjoint. For example, there might be a student majoring in both math and physics. Such a student would be counted

[4] See *The Best Card Trick* by Michael Kleber for more information.

twice on the right side of this equation, once as an element of $M$ and once as an element of $P$. Worse, there might be a triple-major[5] counted *three* times on the right side!

Our most-complicated counting rule determines the size of a union of sets that are not necessarily disjoint. Before we state the rule, let's build some intuition by considering some easier special cases: unions of just two or three sets.

### 15.12.1 Union of Two Sets

For two sets, $S_1$ and $S_2$, the *Inclusion-Exclusion Rule* is that the size of their union is:

$$|S_1 \cup S_2| = |S_1| + |S_2| - |S_1 \cap S_2| \tag{15.5}$$

Intuitively, each element of $S_1$ is accounted for in the first term, and each element of $S_2$ is accounted for in the second term. Elements in *both* $S_1$ and $S_2$ are counted *twice* —once in the first term and once in the second. This double-counting is corrected by the final term.

### 15.12.2 Union of Three Sets

So how many students are there in the math, EECS, and physics departments? In other words, what is $|M \cup E \cup P|$ if:

$$\begin{aligned} |M| &= 60 \\ |E| &= 200 \\ |P| &= 40. \end{aligned}$$

The size of a union of three sets is given by a more complicated Inclusion-Exclusion formula:

$$\begin{aligned} |S_1 \cup S_2 \cup S_3| = {} & |S_1| + |S_2| + |S_3| \\ & - |S_1 \cap S_2| - |S_1 \cap S_3| - |S_2 \cap S_3| \\ & + |S_1 \cap S_2 \cap S_3|. \end{aligned}$$

Remarkably, the expression on the right accounts for each element in the union of $S_1$, $S_2$, and $S_3$ exactly once. For example, suppose that $x$ is an element of all three sets. Then $x$ is counted three times (by the $|S_1|$, $|S_2|$, and $|S_3|$ terms), subtracted off three times (by the $|S_1 \cap S_2|$, $|S_1 \cap S_3|$, and $|S_2 \cap S_3|$ terms), and then counted once more (by the $|S_1 \cap S_2 \cap S_3|$ term). The net effect is that $x$ is counted just once.

[5] . . . though not at MIT anymore.

If $x$ is in two sets (say, $S_1$ and $S_2$), then $x$ is counted twice (by the $|S_1|$ and $|S_2|$ terms) and subtracted once (by the $|S_1 \cap S_2|$ term). In this case, $x$ does not contribute to any of the other terms, since $x \notin S_3$.

So we can't answer the original question without knowing the sizes of the various intersections. Let's suppose that there are:

| | |
|---|---|
| 4 | math - EECS double majors |
| 3 | math - physics double majors |
| 11 | EECS - physics double majors |
| 2 | triple majors |

Then $|M \cap E| = 4+2$, $|M \cap P| = 3+2$, $|E \cap P| = 11+2$, and $|M \cap E \cap P| = 2$. Plugging all this into the formula gives:

$$\begin{aligned} |M \cup E \cup P| &= |M| + |E| + |P| - |M \cap E| - |M \cap P| - |E \cap P| + |M \cap E \cap P| \\ &= 60 + 200 + 40 - 6 - 5 - 13 + 2 \\ &= 278 \end{aligned}$$

### 15.12.3 Sequences with 42, 04, or 60

In how many permutations of the set $\{0, 1, 2, \ldots, 9\}$ do either 4 and 2, 0 and 4, or 6 and 0 appear consecutively? For example, none of these pairs appears in:

$$(7, 2, 9, 5, 4, 1, 3, 8, 0, 6).$$

The 06 at the end doesn't count; we need 60. On the other hand, both 04 and 60 appear consecutively in this permutation:

$$(7, 2, 5, \underline{6}, \underline{0}, \underline{4}, 3, 8, 1, 9).$$

Let $P_{42}$ be the set of all permutations in which 42 appears. Define $P_{60}$ and $P_{04}$ similarly. Thus, for example, the permutation above is contained in both $P_{60}$ and $P_{04}$, but not $P_{42}$. In these terms, we're looking for the size of the set $P_{42} \cup P_{04} \cup P_{60}$.

First, we must determine the sizes of the individual sets, such as $P_{60}$. We can use a trick: group the 6 and 0 together as a single symbol. Then there is an immediate bijection between permutations of $\{0, 1, 2, \ldots 9\}$ containing 6 and 0 consecutively and permutations of:

$$\{60, 1, 2, 3, 4, 5, 7, 8, 9\}.$$

For example, the following two sequences correspond:

$$(7, 2, 5, \underline{6}, \underline{0}, 4, 3, 8, 1, 9) \longleftrightarrow (7, 2, 5, \underline{60}, 4, 3, 8, 1, 9).$$

There are 9! permutations of the set containing 60, so $|P_{60}| = 9!$ by the Bijection Rule. Similarly, $|P_{04}| = |P_{42}| = 9!$ as well.

Next, we must determine the sizes of the two-way intersections, such as $P_{42} \cap P_{60}$. Using the grouping trick again, there is a bijection with permutations of the set:

$$\{42, 60, 1, 3, 5, 7, 8, 9\}.$$

Thus, $|P_{42} \cap P_{60}| = 8!$. Similarly, $|P_{60} \cap P_{04}| = 8!$ by a bijection with the set:

$$\{604, 1, 2, 3, 5, 7, 8, 9\}.$$

And $|P_{42} \cap P_{04}| = 8!$ as well by a similar argument. Finally, note that $|P_{60} \cap P_{04} \cap P_{42}| = 7!$ by a bijection with the set:

$$\{6042, 1, 3, 5, 7, 8, 9\}.$$

Plugging all this into the formula gives:

$$|P_{42} \cup P_{04} \cup P_{60}| = 9! + 9! + 9! - 8! - 8! - 8! + 7!.$$

### 15.12.4 Union of *n* Sets

The size of a union of $n$ sets is given by the following rule.

**Rule 15.12.1** (Inclusion-Exclusion)**.**

$$|S_1 \cup S_2 \cup \cdots \cup S_n| =$$

| | |
|---|---|
| | *the sum of the sizes of the individual sets* |
| minus | *the sizes of all two-way intersections* |
| plus | *the sizes of all three-way intersections* |
| minus | *the sizes of all four-way intersections* |
| plus | *the sizes of all five-way intersections, etc.* |

The formulas for unions of two and three sets are special cases of this general rule.

This way of expressing Inclusion-Exclusion is easy to understand and nearly as precise as expressing it in mathematical symbols, but we'll need the symbolic version below, so let's work on deciphering it now.

We already have a concise notation for the sum of sizes of the individual sets, namely,

$$\sum_{i=1}^{n} |S_i|.$$

A "two-way intersection" is a set of the form $S_i \cap S_j$ for $i \neq j$. We regard $S_j \cap S_i$ as the same two-way intersection as $S_i \cap S_j$, so we can assume that $i < j$. Now we can express the sum of the sizes of the two-way intersections as

$$\sum_{1 \leq i < j \leq n} |S_i \cap S_j|.$$

Similarly, the sum of the sizes of the three-way intersections is

$$\sum_{1 \leq i < j < k \leq n} |S_i \cap S_j \cap S_k|.$$

These sums have alternating signs in the Inclusion-Exclusion formula, with the sum of the $k$-way intersections getting the sign $(-1)^{k-1}$. This finally leads to a symbolic version of the rule:

**Rule** (Inclusion-Exclusion)**.**

$$\begin{aligned}\left|\bigcup_{i=1}^{n} S_i\right| = &\sum_{i=1}^{n} |S_i| \\ &- \sum_{1 \leq i < j \leq n} |S_i \cap S_j| \\ &+ \sum_{1 \leq i < j < k \leq n} |S_i \cap S_j \cap S_k| + \cdots \\ &+ (-1)^{n-1} \left|\bigcap_{i=1}^{n} S_i\right|.\end{aligned}$$

While it's often handy express the rule in this way as a sum of sums, it is not necessary to group the terms by how many sets are in the intersections. So another way to state the rule is:

**Rule** (Inclusion-Exclusion-II)**.**

$$\left|\bigcup_{i=1}^{n} S_i\right| = \sum_{\emptyset \neq I \subseteq \{1,\ldots,n\}} (-1)^{|I|+1} \left|\bigcap_{i \in I} S_i\right|$$

A proof of these rules using just highschool algebra is given in Problem 15.47.

### 15.12.5 Computing Euler's Function

As an example, let's use Inclusion-Exclusion to derive an explicit formula (15.6) for Euler's function, $\phi(n)$. By definition, $\phi(n)$ is the number of nonnegative integers less than a positive integer $n$ that are relatively prime to $n$. But the set $S$ of nonnegative integers less than $n$ that are *not* relatively prime to $n$ will be easier to count.

Suppose the prime factorization of $n$ is $p_1^{e_1} \cdots p_m^{e_m}$ for distinct primes $p_i$. This means that the integers in $S$ are precisely the nonnegative integers less than $n$ that are divisible by at least one of the $p_i$'s. Letting $C_a$ be the set of nonnegative integers less than $n$ that are divisible by $a$, we have

$$S = \bigcup_{i=1}^{m} C_{p_i}.$$

We'll be able to find the size of this union using Inclusion-Exclusion because the intersections of the $C_p$'s are easy to count. For example, $C_p \cap C_q \cap C_r$ is the set of nonnegative integers less than $n$ that are divisible by each of $p$, $q$ and $r$. But since the $p, q, r$ are distinct primes, being divisible by each of them is the same as being divisible by their product. Now observe that if $k$ is a positive divisor of $n$, then exactly $n/k$ nonnegative integers less than $n$ are divisible by $k$, namely, $0, k, 2k, \ldots, ((n/k) - 1)k$. So exactly $n/pqr$ nonnegative integers less than $n$ are divisible by all three primes $p$, $q$, $r$. In other words,

$$|C_p \cap C_q \cap C_r| = \frac{n}{pqr}.$$

Reasoning this way about all the intersections among the $C_p$'s and applying Inclusion-Exclusion, we get

$$
\begin{aligned}
|S| &= \left|\bigcup_{i=1}^{m} C_{p_i}\right| \\
&= \sum_{i=1}^{m} |C_{p_i}| - \sum_{1\le i<j\le m} |C_{p_i} \cap C_{p_j}| \\
&\quad + \sum_{1\le i<j<k\le m} |C_{p_i} \cap C_{p_j} \cap C_{p_k}| - \cdots + (-1)^{m-1} \left|\bigcap_{i=1}^{m} C_{p_i}\right| \\
&= \sum_{i=1}^{m} \frac{n}{p_i} - \sum_{1\le i<j\le m} \frac{n}{p_i p_j} \\
&\quad + \sum_{1\le i<j<k\le m} \frac{n}{p_i p_j p_k} - \cdots + (-1)^{m-1} \frac{n}{p_1 p_2 \cdots p_n} \\
&= n\left(\sum_{i=1}^{m} \frac{1}{p_i} - \sum_{1\le i<j\le m} \frac{1}{p_i p_j} + \sum_{1\le i<j<k\le m} \frac{1}{p_i p_j p_k} - \cdots + (-1)^{m-1} \frac{1}{p_1 p_2 \cdots p_n}\right)
\end{aligned}
$$

But $\phi(n) = n - |S|$ by definition, so

$$
\begin{aligned}
\phi(n) &= n\left(1 - \sum_{i=1}^{m} \frac{1}{p_i} + \sum_{1\le i<j\le m} \frac{1}{p_i p_j} - \sum_{1\le i<j<k\le m} \frac{1}{p_i p_j p_k} + \cdots + (-1)^{m} \frac{1}{p_1 p_2 \cdots p_n}\right) \\
&= n \prod_{i=1}^{m} \left(1 - \frac{1}{p_i}\right). \qquad (15.6)
\end{aligned}
$$

Yikes! That was pretty hairy. Are you getting tired of all that nasty algebra? If so, then good news is on the way. In the next section, we will show you how to prove some heavy-duty formulas without using any algebra at all. Just a few words and you are done. No kidding.

## 15.13 Combinatorial Proofs

Suppose you have $n$ different T-shirts, but only want to keep $k$. You could equally well select the $k$ shirts you want to keep or select the complementary set of $n - k$ shirts you want to throw out. Thus, the number of ways to select $k$ shirts from

among $n$ must be equal to the number of ways to select $n-k$ shirts from among $n$. Therefore:

$$\binom{n}{k} = \binom{n}{n-k}.$$

This is easy to prove algebraically, since both sides are equal to:

$$\frac{n!}{k!\,(n-k)!}.$$

But we didn't really have to resort to algebra; we just used counting principles.

Hmmm....

### 15.13.1 Pascal's Identity

Bob, famed Math for Computer Science Teaching Assistant, has decided to try out for the US Olympic boxing team. After all, he's watched all of the *Rocky* movies and spent hours in front of a mirror sneering, "Yo, you wanna piece a' *me*?!" Bob figures that $n$ people (including himself) are competing for spots on the team and only $k$ will be selected. As part of maneuvering for a spot on the team, he needs to work out how many different teams are possible. There are two cases to consider:

- Bob *is* selected for the team, and his $k-1$ teammates are selected from among the other $n-1$ competitors. The number of different teams that can be formed in this way is:
$$\binom{n-1}{k-1}.$$
- Bob is *not* selected for the team, and all $k$ team members are selected from among the other $n-1$ competitors. The number of teams that can be formed this way is:
$$\binom{n-1}{k}.$$

All teams of the first type contain Bob, and no team of the second type does; therefore, the two sets of teams are disjoint. Thus, by the Sum Rule, the total number of possible Olympic boxing teams is:

$$\binom{n-1}{k-1} + \binom{n-1}{k}.$$

Ted, equally-famed Teaching Assistant, thinks Bob isn't so tough and so he might as well also try out. He reasons that $n$ people (including himself) are trying out for $k$ spots. Thus, the number of ways to select the team is simply:

$$\binom{n}{k}.$$

Ted and Bob each correctly counted the number of possible boxing teams. Thus, their answers must be equal. So we know:

**Lemma 15.13.1** (Pascal's Identity)**.**

$$\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}. \tag{15.7}$$

This is called *Pascal's Identity*. And we proved it *without any algebra!* Instead, we relied purely on counting techniques.

### 15.13.2 Giving a Combinatorial Proof

A *combinatorial proof* is an argument that establishes an algebraic fact by relying on counting principles. Many such proofs follow the same basic outline:

1. Define a set $S$.
2. Show that $|S| = n$ by counting one way.
3. Show that $|S| = m$ by counting another way.
4. Conclude that $n = m$.

In the preceding example, $S$ was the set of all possible Olympic boxing teams. Bob computed

$$|S| = \binom{n-1}{k-1} + \binom{n-1}{k}$$

by counting one way, and Ted computed

$$|S| = \binom{n}{k}$$

by counting another way. Equating these two expressions gave Pascal's Identity.

### Checking a Combinatorial Proof

Combinatorial proofs are based on counting the same thing in different ways. This is fine when you've become practiced at different counting methods, but when in doubt, you can fall back on bijections and sequence counting to check such proofs.

For example, let's take a closer look at the combinatorial proof of Pascal's Identity (15.7). In this case, the set $S$ of things to be counted is the collection of all size-$k$ subsets of integers in the interval $[1, n]$.

Now we've already counted $S$ one way, via the Bookkeeper Rule, and found $|S| = \binom{n}{k}$. The other "way" corresponds to defining a bijection between $S$ and the disjoint union of two sets $A$ and $B$ where,

$$A ::= \{(1, X) \mid X \subseteq [2, n] \text{ AND } |X| = k - 1\}$$
$$B ::= \{(0, Y) \mid Y \subseteq [2, n] \text{ AND } |Y| = k\}.$$

Clearly $A$ and $B$ are disjoint since the pairs in the two sets have different first coordinates, so $|A \cup B| = |A| + |B|$. Also,

$$|A| = \#\text{ specified sets } X = \binom{n-1}{k-1},$$
$$|B| = \#\text{ specified sets } Y = \binom{n-1}{k}.$$

Now finding a bijection $f : (A \cup B) \to S$ will prove the identity (15.7). In particular, we can define

$$f(c) ::= \begin{cases} X \cup \{1\} & \text{if } c = (1, X), \\ Y & \text{if } c = (0, Y). \end{cases}$$

It should be obvious that $f$ is a bijection.

## 15.13.3 A Colorful Combinatorial Proof

The set that gets counted in a combinatorial proof in different ways is usually defined in terms of simple sequences or sets rather than an elaborate story about Teaching Assistants. Here is another colorful example of a combinatorial argument.

**Theorem 15.13.2.**

$$\sum_{r=0}^{n} \binom{n}{r}\binom{2n}{n-r} = \binom{3n}{n}$$

*Proof.* We give a combinatorial proof. Let $S$ be all $n$-card hands that can be dealt from a deck containing $n$ different red cards and $2n$ different black cards. First, note that every $3n$-element set has

$$|S| = \binom{3n}{n}$$

$n$-element subsets.

From another perspective, the number of hands with exactly $r$ red cards is

$$\binom{n}{r}\binom{2n}{n-r}$$

since there are $\binom{n}{r}$ ways to choose the $r$ red cards and $\binom{2n}{n-r}$ ways to choose the $n-r$ black cards. Since the number of red cards can be anywhere from 0 to $n$, the total number of $n$-card hands is:

$$|S| = \sum_{r=0}^{n} \binom{n}{r}\binom{2n}{n-r}.$$

Equating these two expressions for $|S|$ proves the theorem. ■

### Finding a Combinatorial Proof

Combinatorial proofs are almost magical. Theorem 15.13.2 looks pretty scary, but we proved it without any algebraic manipulations at all. The key to constructing a combinatorial proof is choosing the set $S$ properly, which can be tricky. Generally, the simpler side of the equation should provide some guidance. For example, the right side of Theorem 15.13.2 is $\binom{3n}{n}$, which suggests that it will be helpful to choose $S$ to be all $n$-element subsets of some $3n$-element set.

## Problems for Section 15.2

### Practice Problems

**Problem 15.1.**

Alice is thinking of a number between 1 and 1000.

What is the least number of yes/no questions you could ask her and be guaranteed to discover what it is? (Alice always answers truthfully.)

**(a)**

**Problem 15.2.**
In how many different ways is it possible to answer the next chapter's practice problems if:

- the first problem has four *true/false* questions,
- the second problem requires choosing one of four alternatives, and
- the answer to the third problem is an integer $\geq 15$ and $\leq 20$?

**Problem 15.3.**
How many total functions are there from set $A$ to set $B$ if $|A| = 3$ and $|B| = 7$?

**Problem 15.4.**
Consider a 6 element set $X$ with elements $\{x_1, x_2, x_3, x_4, x_5, x_6\}$.

**(a)** How many subsets of $X$ contain $x_1$?

**(b)** How many subsets of $X$ contain $x_2$ and $x_3$ but do not contain $x_6$?

**Class Problems**

**Problem 15.5.**
A license plate consists of either:

- 3 letters followed by 3 digits (standard plate)
- 5 letters (vanity plate)
- 2 characters—letters or numbers (big shot plate)

Let $L$ be the set of all possible license plates.

**(a)** Express $L$ in terms of

$$\mathcal{A} = \{A, B, C, \ldots, Z\}$$
$$\mathcal{D} = \{0, 1, 2, \ldots, 9\}$$

using unions ($\cup$) and set products ($\times$).

**(b)** Compute $|L|$, the number of different license plates, using the sum and product rules.

**Problem 15.6. (a)** How many of the billion numbers in the range from 1 to $10^9$ contain the digit 1? (*Hint:* How many don't?)

**(b)** There are 20 books arranged in a row on a shelf. Describe a bijection between ways of choosing 6 of these books so that no two adjacent books are selected and 15-bit strings with exactly 6 ones.

**Problem 15.7.**

**(a)** Let $\mathcal{S}_{n,k}$ be the possible nonnegative integer solutions to the inequality

$$x_1 + x_2 + \cdots + x_k \leq n. \tag{15.8}$$

That is

$$\mathcal{S}_{n,k} ::= \{(x_1, x_2, \ldots, x_k) \in \mathbb{N}^k \mid (15.8) \text{ is true}\}.$$

Describe a bijection between $\mathcal{S}_{n,k}$ and the set of binary strings with $n$ zeroes and $k$ ones.

**(b)** Let $\mathcal{L}_{n,k}$ be the length $k$ weakly increasing sequences of nonnegative integers $\leq n$. That is

$$\mathcal{L}_{n,k} ::= \{(y_1, y_2, \ldots, y_k) \in \mathbb{N}^k \mid y_1 \leq y_2 \leq \cdots \leq y_k \leq n\}.$$

Describe a bijection between $\mathcal{L}_{n,k}$ and $\mathcal{S}_{n,k}$.

**Problem 15.8.**

An $n$-vertex *numbered tree* is a tree whose vertex set is $\{1, 2, \ldots, n\}$ for some $n > 2$. We define the *code* of the numbered tree to be a sequence of $n - 2$ integers from 1 to $n$ obtained by the following recursive process:[6]

> If there are more than two vertices left, write down the *father* of the largest leaf, delete this *leaf*, and continue this process on the resulting smaller tree. If there are only two vertices left, then stop —the code is complete.

For example, the codes of a couple of numbered trees are shown in the Figure 15.7.

[6]The necessarily unique node adjacent to a leaf is called its *father*.

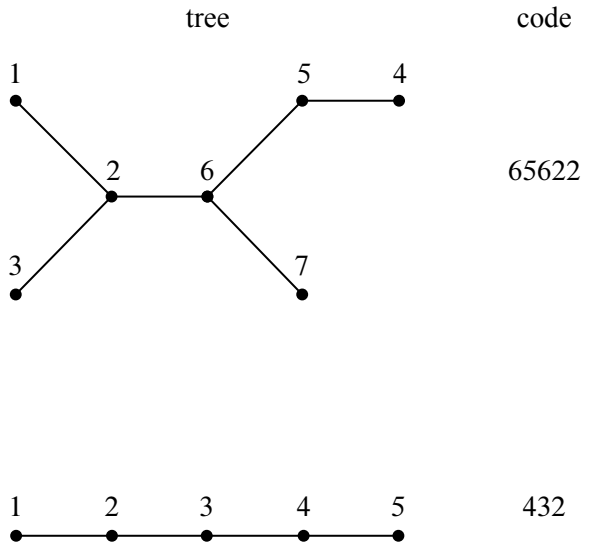



**Figure 15.7**

**(a)** Describe a procedure for reconstructing a numbered tree from its code.

**(b)** Conclude there is a bijection between the $n$-vertex numbered trees and $\{1, \ldots, n\}^{n-2}$, and state how many $n$-vertex numbered trees there are.

**Problem 15.9.**
Let $X$ and $Y$ be finite sets.

**(a)** How many binary relations from $X$ to $Y$ are there?

**(b)** Define a bijection between the set $[X \to Y]$ of all total functions from $X$ to $Y$ and the set $Y^{|X|}$. (Recall $Y^n$ is the cartesian product of $Y$ with itself $n$ times.) Based on that, what is $|[X \to Y]|$?

**(c)** Using the previous part how many *functions*, not necessarily total, are there from $X$ to $Y$? How does the fraction of functions vs. total functions grow as the size of $X$ grows? Is it $O(1)$, $O(|X|)$, $O(2^{|X|})$,...?

**(d)** Show a bijection between the powerset, $\mathcal{P}(X)$, and the set $[X \to \{0,1\}]$ of 0-1-valued total functions on $X$.

**(e)** Let $X ::= \{1, 2, \ldots, n\}$. In this problem we count how many bijections there are from $X$ to itself. Consider the set $B_{X,X}$ of all *bijections* from set $X$ to set $X$.

Show a bijection from $B_{X,X}$ to the set of all permuations of $X$ (as defined in the notes). Using that, count $B_{X,X}$.

## Problems for Section 15.4

### Homework Problems

**Problem 15.10.**
Here is a purely combinatorial proof of Fermat's Little Theorem 8.6.4.

**(a)** Suppose there are beads available in $a$ different colors for some integer $a > 1$, and let $p$ be a prime number. How many different colored length $p$ sequences of beads can be strung together? How many of them contain beads of at least two different colors?

**(b)** Make each string of $p$ beads with at least two colors into a bracelet by tying the two ends of the string together. Two bracelets are the same if one can be rotated to yield the other. (Note, however, that you **cannot** "flip" a bracelet over or reflect it.) Show that for every bracelet, there are exactly $p$ strings of beads that yield it.

*Hint:* Both the fact that $p$ is prime and that the bracelet consists of at least two colors are needed for this to be true.

**(c)** Conclude that $p \mid (a^p - a)$ and from this conclude Fermat's Little Theorem.

## Problems for Section 15.5

### Practice Problems

**Problem 15.11.**
8 students—Anna, Brian, Caine,...–are to be seated around a circular table in a circular room. Two seatings are regarded as defining the same *arrangement* if each student has the same student on his or her right in both seatings: it does not matter which way they face. We'll be interested in counting how many arrangements there are of these 8 students, given some restrictions.

**(a)** As a start, how many different arrangements of these 8 students around the table are there without any restrictions?

**(b)** How many arrangements of these 8 students are there with Anna sitting next to Brian?

**(c)** How many arrangements are there with if Brian sitting next to both Anna AND Caine?

**(d)** How many arrangements are there with Brian sitting next to Anna OR Caine?

**Problem 15.12.**
How many different ways are there to select three dozen colored roses if red, yellow, pink, white, purple and orange roses are available?

**Problem 15.13.**
Suppose you want to select $k$ out of $n$ books on a shelf so that there are always at least 3 unselected books between selected books. Describe a bijection between book selection and bit-strings of length $L$ containing exactly $M$ `1`'s, so that counting the number of all such bit-strings gives us the number of book selections. Find $L$ and $M$ and briefly explain why it works.

(Assume $n$ is large enough for this to be possible.)

**Class Problems**

**Problem 15.14.**
Your class tutorial has 12 students, who are supposed to break up into 4 groups of 3 students each. Your Teaching Assistant (TA) has observed that the students waste too much time trying to form balanced groups, so he decided to pre-assign students to groups and email the group assignments to his students.

**(a)** Your TA has a list of the 12 students in front of him, so he divides the list into consecutive groups of 3. For example, if the list is ABCDEFGHIJKL, the TA would define a sequence of four groups to be $(\{A, B, C\}, \{D, E, F\}, \{G, H, I\}, \{J, K, L\})$. This way of forming groups defines a mapping from a list of twelve students to a sequence of four groups. This is a $k$-to-1 mapping for what $k$?

**(b)** A group assignment specifies which students are in the same group, but not any order in which the groups should be listed. If we map a sequence of 4 groups,

$$(\{A, B, C\}, \{D, E, F\}, \{G, H, I\}, \{J, K, L\}),$$

into a group assignment

$$\{\{A, B, C\}, \{D, E, F\}, \{G, H, I\}, \{J, K, L\}\},$$

this mapping is $j$-to-1 for what $j$?

**(c)** How many group assignments are possible?

**(d)** In how many ways can $3n$ students be broken up into $n$ groups of 3?

**Problem 15.15.**
A pizza house is having a promotional sale. Their commercial reads:

> We offer 9 different toppings for your pizza! Buy 3 large pizzas at the regular price, and you can get each one with as many different toppings as you wish, absolutely free. That's 22,369,621 different ways to choose your pizzas!

The ad writer was a former Harvard student who had evaluated the formula $(2^9)^3/3!$ on his calculator and gotten close to 22,369,621. Unfortunately, $(2^9)^3/3!$ is obviously not an integer, so clearly something is wrong. What mistaken reasoning might have led the ad writer to this formula? Explain how to fix the mistake and get a correct formula.

**Problem 15.16.**
Answer the following quesions using the Generalized Product Rule.

**(a)** Next week, I'm going to get really fit! On day 1, I'll exercise for 5 minutes. On each subsequent day, I'll exercise 0, 1, 2, or 3 minutes more than the previous day. For example, the number of minutes that I exercise on the seven days of next week might be 5, 6, 9, 9, 9, 11, 12. How many such sequences are possible?

**(b)** An *r-permutation* of a set is a sequence of $r$ distinct elements of that set. For example, here are all the 2-permutations of $\{a, b, c, d\}$:

$$\begin{array}{lll} (a,b) & (a,c) & (a,d) \\ (b,a) & (b,c) & (b,d) \\ (c,a) & (c,b) & (c,d) \\ (d,a) & (d,b) & (d,c) \end{array}$$

How many $r$-permutations of an $n$-element set are there? Express your answer using factorial notation.

**(c)** How many $n \times n$ matrices are there with *distinct* entries drawn from $\{1, \ldots, p\}$, where $p \geq n^2$?

**Problem 15.17. (a)** There are 30 books arranged in a row on a shelf. In how many ways can eight of these books be selected so that there are at least two unselected books between any two selected books?

**(b)** How many nonnegative integer solutions are there for the following equality?

$$x_1 + x_2 + \cdots + x_m = k. \tag{15.9}$$

**(c)** How many nonnegative integer solutions are there for the following inequality?

$$x_1 + x_2 + \cdots + x_m \leq k. \tag{15.10}$$

**(d)** How many length-$m$ weakly increasing sequences of nonnegative integers $\leq k$ are there?

### Homework Problems

**Problem 15.18.**
This problem is about binary relations on the set of integers in the interval $[1, n]$, and digraphs and simple graphs whose vertex set is $[1, n]$.

**(a)** How many digraphs are there?

**(b)** How many simple graphs are there?

**(c)** How many asymmetric binary relations are there?

**(d)** How many path-total strict partial orders are there?

**Problem 15.19.**
Answer the following questions with a number or a simple formula involving factorials and binomial coefficients. Briefly explain your answers.

**(a)** How many ways are there to order the 26 letters of the alphabet so that no two of the vowels `a`, `e`, `i`, `o`, `u` appear consecutively and the last letter in the ordering is not a vowel?

*Hint:* Every vowel appears to the left of a consonant.

**(b)** How many ways are there to order the 26 letters of the alphabet so that there are *at least two* consonants immediately following each vowel?

**(c)** In how many different ways can $2n$ students be paired up?

**(d)** Two $n$-digit sequences of digits 0,1,...,9 are said to be of the *same type* if the digits of one are a permutation of the digits of the other. For $n = 8$, for example, the sequences `03088929` and `00238899` are the same type. How many types of $n$-digit integers are there?

**Problem 15.20.**
In a standard 52-card deck, each card has one of thirteen *rank*s in the set, $R$, and one of four *suit*s in the set, $S$, where

$$R ::= \{A, 2, \ldots, 10, J, Q, K\},$$
$$S ::= \{\clubsuit, \diamondsuit, \heartsuit, \spadesuit\}.$$

A 5-card *hand* is a set of five distinct cards from the deck.

For each part describe a bijection between a set that can easily be counted using the Product and Sum Rules of Ch. 15.1, and the set of hands matching the specification. *Give bijections, not numerical answers.*

For instance, consider the set of 5-card hands containing all 4 suits. Each such hand must have 2 cards of one suit. We can describe a bijection between such hands and the set $S \times R_2 \times R^3$ where $R_2$ is the set of two-element subsets of $R$. Namely, an element

$$(s, \{r_1, r_2\}, (r_3, r_4, r_5)) \in S \times R_2 \times R^3$$

indicates

1. the repeated suit, $s \in S$,
2. the set, $\{r_1, r_2\} \in R_2$, of ranks of the cards of suit, $s$, and
3. the ranks $(r_3, r_4, r_5)$ of remaining three cards, listed in increasing suit order where $\clubsuit \prec \diamondsuit \prec \heartsuit \prec \spadesuit$.

For example,

$$(\clubsuit, \{10, A\}, (J, J, 2)) \longleftrightarrow \{A\clubsuit, 10\clubsuit, J\diamondsuit, J\heartsuit, 2\spadesuit\}.$$

**(a)** A single pair of the same rank (no 3-of-a-kind, 4-of-a-kind, or second pair).

**(b)** Three or more aces.

**Problem 15.21.**
Suppose you have seven dice —each a different color of the rainbow; otherwise the dice are standard, with faces numbered 1 to 6. A *roll* is a sequence specifying a value for each die in rainbow (ROYGBIV) order. For example, one roll is $(3, 1, 6, 1, 4, 5, 2)$ indicating that the red die showed a 3, the orange die showed 1, the yellow 6,....

For the problems below, describe a bijection between the specified set of rolls and another set that is easily counted using the Product, Generalized Product, and

similar rules. Then write a simple arithmetic formula, possibly involving factorials and binomial coefficients, for the size of the set of rolls. You do not need to prove that the correspondence between sets you describe is a bijection, and you do not need to simplify the expression you come up with.

For example, let $A$ be the set of rolls where 4 dice come up showing the same number, and the other 3 dice also come up the same, but with a different number. Let $R$ be the set of seven rainbow colors and $S ::= [1, 6]$ be the set of dice values.

Define $B ::= P_{S,2} \times R_3$, where $P_{S,2}$ is the set of 2-permutations of $S$ and $R_3$ is the set of size-3 subsets of $R$. Then define a bijection from $A$ to $B$ by mapping a roll in $A$ to the sequence in $B$ whose first element is an ordered pair consisting of the number that came up three times followed by the number that came up four times, and whose second element is the set of colors of the three matching dice.

For example, the roll

$$(4, 4, 2, 2, 4, 2, 4) \in A$$

maps to

$$((2, 4), \{\text{yellow,green,indigo}\}) \in B.$$

Now by the Bijection rule $|A| = |B|$, and by the Generalized Product and Subset rules,

$$|B| = 6 \cdot 5 \cdot \binom{7}{3}.$$

**(a)** For how many rolls do *exactly* two dice have the value 6 and the remaining five dice all have different values?

Example: $(6, 2, 6, 1, 3, 4, 5)$ is a roll of this type, but $(1, 1, 2, 6, 3, 4, 5)$ and $(6, 6, 1, 2, 4, 3, 4)$ are not.

**(b)** For how many rolls do two dice have the same value and the remaining five dice all have different values?

Example: $(4, 2, 4, 1, 3, 6, 5)$ is a roll of this type, but $(1, 1, 2, 6, 1, 4, 5)$ and $(6, 6, 1, 2, 4, 3, 4)$ are not.

**(c)** For how many rolls do two dice have one value, two different dice have a second value, and the remaining three dice a third value?

Example: $(6, 1, 2, 1, 2, 6, 6)$ is a roll of this type, but $(4, 4, 4, 4, 1, 3, 5)$ and $(5, 5, 5, 6, 6, 1, 2)$ are not.

### Exam Problems

**Problem 15.22.**
Suppose that two identical 52-card decks are mixed together. Write a simple formula for the number of distinct permutations of the 104 cards.

## Problems for Section 15.6

### Practice Problems

**Problem 15.23.**
How many different permutations are there of the sequence of letters in "MISSISSIPPI"?

### Exam Problems

**Problem 15.24.**
There is a robot that steps between integer positions in 3-dimensional space. Each step of the robot increments one coordinate and leaves the other two unchanged.

**(a)** How many paths can the robot follow going from the origin $(0, 0, 0)$ to $(3, 4, 5)$?

**(b)** How many paths can the robot follow going from the origin $(i, j, k)$ to $(m, n, p)$?

## Problems for Section 15.7

### Practice Problems

**Problem 15.25.**
Find the coefficients of $x^{10}y^5$ in $(19x + 4y)^{15}$

**Problem 15.26.**
Find the coefficient of $x^4$ in the following expressions.

**(a)** $(x + 1)^9$?

**(b)** $(3x + 2)^6$ ?

### Class Problems

**Problem 15.27.**
Find the coefficients of

**(a)** $x^5$ in $(1 + x)^{11}$

**(b)** $x^8y^9$ in $(3x + 2y)^{17}$

**(c)** $a^6b^6$ in $(a^2+b^3)^5$

**Problem 15.28. (a)** Use the Multinomial Theorem 15.7.2 to prove that

$$(x_1 + x_2 + \cdots + x_n)^p \equiv x_1^p + x_2^p + \cdots + x_n^p \pmod{p} \tag{15.11}$$

for all primes $p$. (Do not prove it using Fermat's "little" Theorem. The point of this problem is to offer an independent proof of Fermat's theorem.)

*Hint:* Explain why $\binom{p}{k_1,k_2,\ldots,k_n}$ is divisible by $p$ if all the $k_i$'s are positive integers less than $p$.

**(b)** Explain how (15.11) immediately proves Fermat's Little Theorem 8.6.4: $n^{p-1} \equiv 1 \pmod{p}$ when $n$ is not a multiple of $p$.

## Homework Problems

**Problem 15.29.**

The *degree sequence* of a simple graph is the weakly decreasing sequence of degrees of its vertices. For example, the degree sequence for the 5-vertex numbered tree pictured in the Figure 15.7 in Problem 15.8 is $(2, 2, 2, 1, 1)$ and for the 7-vertex tree it is $(3, 3, 2, 1, 1, 1, 1)$.

We're interested in counting how many numbered trees there are with a given degree sequence. We'll do this using the bijection defined in Problem 15.8 between $n$-vertex numbered trees and length $n-2$ code words whose characters are integers between 1 and $n$.

The *occurrence number* for a character in a word is the number of times that the character occurs in the word. For example, in the word `65622`, the occurrence number for `6` is two, and the occurrence number for `5` is one. The *occurrence sequence* of a word is the weakly decreasing sequence of occurrence numbers of characters in the word. The occurrence sequence for this word is $(2, 2, 1)$ because it has two occurrences of each of the characters `6` and `2`, and one occurrence of `5`.

**(a)** There is simple relationship between the degree sequence of an $n$-vertex numbered tree and the occurrence sequence of its code. Describe this relationship and explain why it holds. Conclude that counting $n$-vertex numbered trees with a given degree sequence is the same as counting the number of length $n-2$ code words with a given occurrence sequence.

*Hint:* How many times does a vertex of degree, $d$, occur in the code?

For simplicity, let's focus on counting 9-vertex numbered trees with a given degree sequence. By part (a), this is the same as counting the number of length 7 code words with a given occurrence sequence.

Any length 7 code word has a *pattern*, which is another length 7 word over the alphabet a,b,c,d,e,f,g that has the same occurrence sequence.

**(b)** How many length 7 patterns are there with three occurrences of a, two occurrences of b, and one occurrence of c and d?

**(c)** How many ways are there to assign occurrence numbers to integers $1, 2, \ldots, 9$ so that a code word with those occurrence numbers would have the occurrence sequence $3, 2, 1, 1, 0, 0, 0, 0, 0$?

In general, to find the pattern of a code word, list its characters in decreasing order by *number of occurrences*, and list characters with the same number of occurrences in decreasing order. Then replace successive characters in the list by successive letters a,b,c,d,e,f,g. The code word 2468751, for example, has the pattern fecabdg, which is obtained by replacing its characters 8,7,6,5,4,2,1 by a,b,c,d,e,f,g, respectively. The code word 2449249 has pattern caabcab, which is obtained by replacing its characters 4,9,2 by a,b,c, respectively.

**(d)** What length 7 code word has three occurrences of 7, two occurrences of 8, one occurrence each of 2 and 9, and pattern abacbad?

**(e)** Explain why the number of 9-vertex numbered trees with degree sequence $(4, 3, 2, 2, 1, 1, 1, 1, 1)$ is the product of the answers to parts (b) and (c).

## Problems for Section 15.8

### Class Problems

**Problem 15.30.**
The Tao of BOOKKEEPER: we seek enlightenment through contemplation of the word $BOOKKEEPER$.

**(a)** In how many ways can you arrange the letters in the word $POKE$?

**(b)** In how many ways can you arrange the letters in the word $BO_1O_2K$? Observe that we have subscripted the O's to make them distinct symbols.

**(c)** Suppose we map arrangements of the letters in $BO_1O_2K$ to arrangements of the letters in $BOOK$ by erasing the subscripts. Indicate with arrows how the arrangements on the left are mapped to the arrangements on the right.

$$\begin{array}{l} O_2BO_1K \\ KO_2BO_1 \\ O_1BO_2K \\ KO_1BO_2 \\ BO_1O_2K \\ BO_2O_1K \\ \ldots \end{array} \qquad \begin{array}{l} BOOK \\ OBOK \\ KOBO \\ \ldots \end{array}$$

**(d)** What kind of mapping is this, young grasshopper?

**(e)** In light of the Division Rule, how many arrangements are there of $BOOK$?

**(f)** Very good, young master! How many arrangements are there of the letters in $KE_1E_2PE_3R$?

**(g)** Suppose we map each arrangement of $KE_1E_2PE_3R$ to an arrangement of $KEEPER$ by erasing subscripts. List all the different arrangements of $KE_1E_2PE_3R$ that are mapped to $REPEEK$ in this way.

**(h)** What kind of mapping is this?

**(i)** So how many arrangements are there of the letters in $KEEPER$?

*Now you are ready to face the BOOKKEEPER!*

**(j)** How many arrangements of $BO_1O_2K_1K_2E_1E_2PE_3R$ are there?

**(k)** How many arrangements of $BOOK_1K_2E_1E_2PE_3R$ are there?

**(l)** How many arrangements of $BOOKKE_1E_2PE_3R$ are there?

**(m)** How many arrangements of $BOOKKEEPER$ are there?

*Remember well what you have learned: subscripts on, subscripts off.*
*This is the Tao of Bookkeeper.*

**(n)** How many arrangements of $VOODOODOLL$ are there?

**(o)** How many length 52 sequences of digits contain exactly 17 two's, 23 fives, and 12 nines?

## Problems for Section 15.9

### Practice Problems

**Problem 15.31.**
Indicate how many 5-card hands there are of each of the following kinds.

**(a)** A **Sequence** is a hand consisting of five consecutive cards of any suit, such as

$$5\heartsuit - 6\heartsuit - 7\spadesuit - 8\diamondsuit - 9\clubsuit.$$

Note that an Ace may either be high (as in 10-J-Q-K-A), or low (as in A-2-3-4-5), but can't go "around the corner" (that is, Q-K-A-2-3 is *not* a sequence).

How many different **Sequence** hands are possible?

**(b)** A **Matching Suit** is a hand consisting of cards that are all of the same suit in any order.

How many different **Matching Suit** hands are possible?

**(c)** A **Straight Flush** is a hand that is both a *Sequence* and a *Matching Suit*.

How many different **Straight Flush** hands are possible?

**(d)** A **Straight** is a hand that is a *Sequence* but not a *Matching Suit*.

How many possible **Straights** are there?

**(e)** A **Flush** is a hand that is a *Matching Suit* but not a *Sequence*.

How many possible **Flushes** are there?

### Class Problems

**Problem 15.32.**
Solve the following counting problems. Define an appropriate mapping (bijective or $k$-to-1) between a set whose size you know and the set in question.

**(a)** An independent living group is hosting nine new candidates for membership. Each candidate must be assigned a task: 1 must wash pots, 2 must clean the kitchen, 3 must clean the bathrooms, 1 must clean the common area, and 2 must serve dinner. Write a multinomial coefficient for the number of ways this can be done.

**(b)** How many nonnegative integers less than 1,000,000 have exactly one digit equal to 9 and have a sum of digits equal to 17?

## Exam Problems

**Problem 15.33.**
Here are the solutions to the next 10 short answer questions, in no particular order. Enter the solution number after each question.

$$1.\ \frac{n!}{(n-m)!} \qquad 2.\ \binom{n+m}{m} \qquad 3.\ (n-m)! \qquad 4.\ m^n$$

$$5.\ \binom{n-1+m}{m} \qquad 6.\ \binom{n-1+m}{n} \qquad 7.\ 2^{mn} \qquad 8.\ n^m$$

**(a)** How many solutions over the nonnegative integers are there to the inequality

$$x_1 + x_2 + \cdots + x_n \leq m\ ?$$

**(b)** How many length $m$ words can be formed from an $n$-letter alphabet, if no letter is used more than once?

**(c)** How many length $m$ words can be formed from an $n$-letter alphabet, if letters can be reused?

**(d)** How many binary relations are there from set $A$ to set $B$ when $|A| = m$ and $|B| = n$?

**(e)** How many total injective functions are there from set $A$ to set $B$, where $|A| = m$ and $|B| = n \geq m$?

**(f)** How many ways are there to place a total of $m$ distinguishable balls into $n$ distinguishable urns, with some urns possibly empty or with several balls?

**(g)** How many ways are there to place a total of $m$ indistinguishable balls into $n$ distinguishable urns, with some urns possibly empty or with several balls?

**(h)** How many ways are there to put a total of $m$ distinguishable balls into $n$ distinguishable urns with at most one ball in each urn?

**Problem 15.34. (a)** How many solutions over the *positive* integers are there to the inequality:

$$x_1 + x_2 + \ldots + x_{10} \leq 100$$

**(b)** In how many ways can Mr. and Mrs. Grumperson distribute 13 identical pieces of coal to their three children for Christmas so that each child gets at least one piece?

## Problems for Section 15.10

### Practice Problems

**Problem 15.35.**
Below is a list of properties that a group of people might possess.

For each property, either give the minimum number of people that must be in a group to ensure that the property holds, or else indicate that the property need not hold even for arbitrarily large groups of people.

(Assume that every year has exactly 365 days; ignore leap years.)

**(a)** At least 2 people were born on the same day of the year (ignore year of birth).

**(b)** At least 2 people were born on January 1.

**(c)** At least 3 people were born on the same day of the week.

**(d)** At least 4 people were born in the same month.

**(e)** At least 2 people were born exactly one week apart.

### Class Problems

**Problem 15.36.**
Solve the following problems using the pigeonhole principle. For each problem, try to identify the *pigeons*, the *pigeonholes*, and a *rule* assigning each pigeon to a pigeonhole.

**(a)** In a certain Institute of Technology, Every ID number starts with a 9. Suppose that each of the 75 students in a class sums the nine digits of their ID number. Explain why two people must arrive at the same sum.

**(b)** In every set of 100 integers, there exist two whose difference is a multiple of 37.

**(c)** For any five points inside a unit square (not on the boundary), there are two points at distance *less than* $1/\sqrt{2}$.

**(d)** Show that if $n + 1$ numbers are selected from $\{1, 2, 3, \ldots, 2n\}$, two must be consecutive, that is, equal to $k$ and $k + 1$ for some $k$.

### Homework Problems

**Problem 15.37.**
**Pigeon Huntin'**

**(a)** Show that any odd integer $x$ in the range $10^9 < x < 2 \cdot 10^9$ containing all ten digits $0, 1, \ldots, 9$ must have consecutive even digits. *Hint:* What can you conclude about the parities of the first and last digit?

**(b)** Show that there are 2 vertices of equal degree in any finite undirected graph with $n \geq 2$ vertices. *Hint:* Cases conditioned upon the existence of a degree zero vertex.

**Problem 15.38.**
Show that for any set of 201 positive integers less than 300, there must be two whose quotient is a power of three (with no remainder).

**Problem 15.39. (a)** Color each point in the plane with integer coordinates either red, white or blue. Let $R$ be a $4 \times 82$ rectangular grid of these points. Explain why at least two of the 82 rows in $R$ must have the same sequence colors.

**(b)** Conclude that $R$ contains four points with the same color that form the corners of a rectangle.

**(c)** Generalize the above argument to a coloring using the rainbow colors Red, Orange, Yellow, Green, Blue, Indigo, Violet as well as White and Black.

## Problems for Section 15.11

### Class Problems

**Problem 15.40. (a)** Show that the Magician could not pull off the trick with a deck larger than 124 cards.

*Hint:* Compare the number of 5-card hands in an $n$-card deck with the number of 4-card sequences.

**(b)** Show that, in principle, the Magician could pull off the Card Trick with a deck of 124 cards.

*Hint:* Hall's Theorem and degree-constrained (11.5.5) graphs.

**Problem 15.41.**
The Magician can determine the 5th card in a poker hand when his Assisant reveals the other 4 cards. Describe a similar method for determining 2 hidden cards in a hand of 9 cards when your Assisant reveals the other 7 cards.

### Homework Problems

**Problem 15.42.**
Section 15.11.3 explained why it is not possible to perform a four-card variant of the hidden-card magic trick with one card hidden. But the Magician and her Assistant are determined to find a way to make a trick like this work. They decide to change the rules slightly: instead of the Assistant lining up the three unhidden cards for the Magician to see, he will line up all four cards with one card face down and the other three visible. We'll call this the *face-down four-card trick*.

For example, suppose the audience members had selected the cards 9♡, 10♢, *A*♣, 5♣. Then the Assistant could choose to arrange the 4 cards in any order so long as one is face down and the others are visible. Two possibilities are:

| | | | |
|---|---|---|---|
| *A*♣ | ? | 10♢ | 5♣ |
| ? | 5♣ | 9♡ | 10♢ |

**(a)** Explain how to model this the face-down four-card trick as a matching problem, and show that there must be a bipartite matching which theoretically will allow the Magician and Assistant to perform the trick.

**(b)** There is actually a simple way to perform the face-down four-card trick.[7]

[7] This elegant method was devised in Fall '09 by student Katie E Everett.

**Case 1**. *there are two cards with the same suit*: Say there are two ♠ cards. The Assistant proceeds as in the original card trick: he puts one of the ♠ cards *face up as the first card*. He will place the second ♠ card *face down*. He then uses a permutation of the face down card and the remaining two face up cards to code the offset of the face down card from the first card.

**Case 2**. *all four cards have different suits*: Assign numbers $0, 1, 2, 3$ to the four suits in some agreed upon way. The Assistant computes, $s$, the sum modulo 4 of the ranks of the four cards, and chooses the card with suit $s$ to be placed *face down as the first card*. He then uses a permutation of the remaining three face-up cards to code the rank of the face down card.

Explain how in Case 2. the Magician can determine the face down card from the cards the Assistant shows her.

**(c)** Explain how any method for performing the face-down four-card trick can be adapted to perform the regular (5-card hand, show 4 cards) with a 52-card deck consisting of the usual 52 cards along with a 53rd card call the *joker*.

## Problems for Section 15.12

### Practice Problems

**Problem 15.43.**
Let $A_1$, $A_2$, $A_3$ be sets with $|A_1| = 100$, $|A_2| = 1,000$, and $|A_3| = 10,000$.
Determine $|A_1 \cup A_2 \cup A_3|$ in each of the following cases:

**(a)** $A_1 \subset A_2 \subset A_3$.

**(b)** The sets are pairwise disjoint.

**(c)** For any two of the sets, there is exactly one element in both.

**(d)** There are two elements common to each pair of sets and one element in all three sets.

**Problem 15.44.**
The working days in the next year can be numbered 1, 2, 3, ..., 300. I'd like to avoid as many as possible.

- On even-numbered days, I'll say I'm sick.
- On days that are a multiple of 3, I'll say I was stuck in traffic.

- On days that are a multiple of 5, I'll refuse to come out from under the blankets.

In total, how many work days will I *avoid* in the coming year?

### Class Problems

**Problem 15.45.**
A certain company wants to have security for their computer systems. So they have given everyone a password. A length 10 word containing each of the characters:

a, d, e, f, i, l, o, p, r, s,

is called a *cword*. A password will be a cword which does not contain any of the subwords "fails", "failed", or "drop".

For example, the following two words are passwords: adefiloprs, srpolifeda, but the following three cwords are not: a**drop**eflis, **failedrop**s, **drop**e**fails**.

**(a)** How many cwords contain the subword "drop"?

**(b)** How many cwords contain both "drop" and "fails"?

**(c)** Use the Inclusion-Exclusion Principle to find a simple arithmetic formula involving factorials for the number of passwords.

**Problem 15.46.**
We want to count step-by-step paths between points in the plane with integer coordinates. Ony two kinds of step are allowed: a right-step which increments the $x$ coordinate, and an up-step which increments the $y$ coordinate.

**(a)** How many paths are there from $(0, 0)$ to $(20, 30)$?

**(b)** How many paths are there from $(0, 0)$ to $(20, 30)$ that go through the point $(10, 10)$?

**(c)** How many paths are there from $(0, 0)$ to $(20, 30)$ that do *not* go through either of the points $(10, 10)$ and $(15, 20)$?

*Hint:* Let $P$ be the set of paths from $(0, 0)$ to $(20, 30)$, $N_1$ be the paths in $P$ that go through $(10, 10)$ and $N_2$ be the paths in $P$ that go through $(15, 20)$.

**Problem 15.47.**
Let's develop a proof of the Inclusion-Exclusion formula using high school algebra.

**(a)** Most high school students will get freaked by the following formula, even though they actually know the rule it expresses. How would you explain it to them?

$$\prod_{i=1}^{n}(1-x_i) = \sum_{I\subseteq\{1,\ldots,n\}}(-1)^{|I|}\prod_{j\in I}x_j. \tag{15.12}$$

*Hint:* Show them an example.

For any set, $S$, let $M_S$ be the *membership* function of $S$:

$$M_S(x) = \begin{cases} 1 & \text{if } x \in S, \\ 0 & \text{if } x \notin S. \end{cases}$$

Let $S_1, \ldots, S_n$ be a sequence of finite sets, and abbreviate $M_{S_i}$ as $M_i$. Let the domain of discourse, $D$, be the union of the $S_i$'s. That is, we let

$$D ::= \bigcup_{i=1}^{n} S_i,$$

and take complements with respect to $D$, that is,

$$\overline{T} ::= D - T,$$

for $T \subseteq D$.

**(b)** Verify that for $T \subseteq D$ and $I \subseteq \{1, \ldots n\}$,

$$M_{\overline{T}} = 1 - M_T, \tag{15.13}$$

$$M_{(\bigcap_{i\in I} S_i)} = \prod_{i\in I} M_{S_i}, \tag{15.14}$$

$$M_{(\bigcup_{i\in I} S_i)} = 1 - \prod_{i\in I}(1-M_i). \tag{15.15}$$

(Note that (15.14) holds when $I$ is empty because, by convention, an empty product equals 1, and an empty intersection equals the domain of discourse, $D$.)

**(c)** Use (15.12) and (15.15) to prove

$$M_D = \sum_{\emptyset\neq I\subseteq\{1,\ldots,n\}}(-1)^{|I|+1}\prod_{j\in I}M_j. \tag{15.16}$$

**(d)** Prove that

$$|T| = \sum_{u \in D} M_T(u). \tag{15.17}$$

**(e)** Now use the previous parts to prove

$$|D| = \sum_{\emptyset \neq I \subseteq \{1,\ldots,n\}} (-1)^{|I|+1} \left| \bigcap_{i \in I} S_i \right| \tag{15.18}$$

**(f)** Finally, explain why (15.18) immediately implies the usual form of the Inclusion-Exclusion Principle:

$$|D| = \sum_{i=1}^{n} (-1)^{i+1} \sum_{\substack{I \subseteq \{1,\ldots,n\} \\ |I|=i}} \left| \bigcap_{j \in I} S_j \right|. \tag{15.19}$$

## Homework Problems

**Problem 15.48.**

How many paths are there from point $(0, 0)$ to $(50, 50)$ if every step increments one coordinate and leaves the other unchanged? How many are there when there are impassable boulders sitting at points $(10, 11)$ and $(21, 20)$? (You do not have to calculate the number explicitly; your answer may be an expression involving binomial coefficients.)

*Hint:* Count the number of paths going through $(10, 11)$, the number through $(21, 20)$, and use Inclusion-Exclusion.

**Problem 15.49.**

A *derangement* is a permutation $(x_1, x_2, \ldots, x_n)$ of the set $\{1, 2, \ldots, n\}$ such that $x_i \neq i$ for all $i$. For example, $(2, 3, 4, 5, 1)$ is a derangement, but $(2, 1, 3, 5, 4)$ is not because 3 appears in the third position. The objective of this problem is to count derangements.

It turns out to be easier to start by counting the permutations that are *not* derangements. Let $S_i$ be the set of all permutations $(x_1, x_2, \ldots, x_n)$ that are not derangements because $x_i = i$. So the set of non-derangements is

$$\bigcup_{i=1}^{n} S_i.$$

**(a)** What is $|S_i|$?

**(b)** What is $\left|S_i \cap S_j\right|$ where $i \neq j$?

**(c)** What is $\left|S_{i_1} \cap S_{i_2} \cap \cdots \cap S_{i_k}\right|$ where $i_1, i_2, \ldots, i_k$ are all distinct?

**(d)** Use the inclusion-exclusion formula to express the number of non-derangements in terms of sizes of possible intersections of the sets $S_1, \ldots, S_n$.

**(e)** How many terms in the expression in part (d) have the form $\left|S_{i_1} \cap S_{i_2} \cap \cdots \cap S_{i_k}\right|$?

**(f)** Combine your answers to the preceding parts to prove the number of non-derangements is:

$$n!\left(\frac{1}{1!} - \frac{1}{2!} + \frac{1}{3!} - \cdots \pm \frac{1}{n!}\right).$$

Conclude that the number of derangements is

$$n!\left(1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \cdots \pm \frac{1}{n!}\right).$$

**(g)** As $n$ goes to infinity, the number of derangements approaches a constant fraction of all permutations. What is that constant? *Hint:*

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots$$

**Problem 15.50.**

How many of the numbers $2, \ldots, n$ are prime? The Inclusion-Exclusion Principle offers a useful way to calculate the answer when $n$ is large. Actually, we will use Inclusion-Exclusion to count the number of *composite* (nonprime) integers from 2 to $n$. Subtracting this from $n - 1$ gives the number of primes.

Let $C_n$ be the set of composites from 2 to $n$, and let $A_m$ be the set of numbers in the range $m + 1, \ldots, n$ that are divisible by $m$. Notice that by definition, $A_m = \emptyset$ for $m \geq n$. So

$$C_n = \bigcup_{i=2}^{n-1} A_i. \tag{15.20}$$

**(a)** Verify that if $m \mid k$, then $A_m \supseteq A_k$.

**(b)** Explain why the right hand side of (15.20) equals

$$\bigcup_{\text{primes } p \leq \sqrt{n}} A_p. \tag{15.21}$$

**(c)** Explain why $|A_m| = \lfloor n/m \rfloor - 1$ for $m \geq 2$.

**(d)** Consider any two relatively prime numbers $p, q \leq n$. What is the one number in $(A_p \cap A_q) - A_{p \cdot q}$?

**(e)** Let $\mathcal{P}$ be a finite set of at least two primes. Give a simple formula for

$$\left| \bigcap_{p \in \mathcal{P}} A_p \right|.$$

**(f)** Use the Inclusion-Exclusion principle to obtain a formula for $|C_{150}|$ in terms the sizes of intersections among the sets $A_2, A_3, A_5, A_7, A_{11}$. (Omit the intersections that are empty; for example, any intersection of more than three of these sets must be empty.)

**(g)** Use this formula to find the number of primes up to 150.

### Exam Problems

**Problem 15.51. (a)** How many length $n$ binary strings are there in which `011` occurs starting at the 4th position?

**(b)** Let $A_i$ be the set of length $n$ binary strings in which `011` occurs starting at the $i$th position. (So $A_i$ is empty for $i > n - 2$.) For $i < j$, the intersections $A_i \cap A_j$ that are nonempty are all the same size. What is $|A_i \cap A_j|$ in this case?

**(c)** Let $t$ be the number of intersections $A_i \cap A_j$ that are nonempty, where $i < j$. Express $t$ as a binomial coefficient.

**(d)** How many length 9 binary strings are there that contain the substring `011`? You should express your answer as an integer or as a simple expression which may include the constant, $t$, of part (c).

*Hint:* Inclusion-exclusion for $|\bigcup_1^7 A_i|$.

**Problem 15.52.**
There are 10 students $A, B, \ldots, J$ who will be lined up left to right according to the some rules below.

Rule I: Student A must not be rightmost.

Rule II: Student B must be adjacent to C (directly to the left or right of C).

Rule III: Student D is always second.

You may answer the following questions with a numerical formula that may involve factorials.

**(a)** How many possible lineups are there that satisfy all three of these rules?

**(b)** How many possible lineups are there that satisfy at least one of these rules?

## Problems for Section 15.13

### Class Problems

**Problem 15.53.**
According to the Multinomial theorem, $(w + x + y + z)^n$ can be expressed as a sum of terms of the form

$$\binom{n}{r_1, r_2, r_3, r_4} w^{r_1} x^{r_2} y^{r_3} z^{r_4}.$$

**(a)** How many terms are there in the sum?

**(b)** The sum of these multinomial coefficients has an easily expressed value. What is it?

$$\sum_{\substack{r_1+r_2+r_3+r_4=n, \\ r_i \in \mathbb{N}}} \binom{n}{r_1,\ r_2,\ r_3,\ r_4} = ? \tag{15.22}$$

*Hint:* How many terms are there when $(w + x + y + z)^n$ is expressed as a sum of monomials in $w, x, y, z$ *before* terms with like powers of these variables are collected together under a single coefficient?

**Problem 15.54.**

**(a)** Give a combinatorial proof of the following identity by letting $S$ be the set of all length-$n$ sequences of letters $a$, $b$ and a single $c$ and counting $|S|$ is two different ways.

$$n2^{n-1} = \sum_{k=1}^{n} k\binom{n}{k} \tag{15.23}$$

**(b)** Now prove (15.23) algebraically by applying the Binomial Theorem to $(1 + x)^n$ and taking derivatives.

**Problem 15.55.**
What do the following expressions equal? Give both algebraic and combinatorial proofs for your answers.

**(a)**

$$\sum_{i=0}^{n} \binom{n}{i}$$

**(b)**

$$\sum_{i=0}^{n} \binom{n}{i}(-1)^i$$

*Hint:* Consider the bit strings with an even number of ones and an odd number of ones.

## Homework Problems

**Problem 15.56.**
Prove the following identity by algebraic manipulation and by giving a combinatorial argument:

$$\binom{n}{r}\binom{r}{k} = \binom{n}{k}\binom{n-k}{r-k}$$

**Problem 15.57. (a)** Find a combinatorial (*not* algebraic) proof that

$$\sum_{i=0}^{n} \binom{n}{i} = 2^n.$$

**(b)** Below is a combinatorial proof of an equation. What is the equation?

*Proof.* Stinky Peterson owns $n$ newts, $t$ toads, and $s$ slugs. Conveniently, he lives in a dorm with $n + t + s$ other students. (The students are distinguishable, but creatures of the same variety are not distinguishable.) Stinky wants to put one creature in each neighbor's bed. Let $W$ be the set of all ways in which this can be done.

On one hand, he could first determine who gets the slugs. Then, he could decide who among his remaining neighbors has earned a toad. Therefore, $|W|$ is equal to the expression on the left.

On the other hand, Stinky could first decide which people deserve newts and slugs and then, from among those, determine who truly merits a newt. This shows that $|W|$ is equal to the expression on the right.

Since both expressions are equal to $|W|$, they must be equal to each other. ■

(Combinatorial proofs are real proofs. They are not only rigorous, but also convey an intuitive understanding that a purely algebraic argument might not reveal. However, combinatorial proofs are usually less colorful than this one.)

**Problem 15.58.**
According to the Multinomial Theorem 15.7.2, $(x_1 + x_2 + \cdots + x_k)^n$ can be expressed as a sum of terms of the form

$$\binom{n}{r_1, r_2, \ldots, r_k} x_1^{r_1} x_2^{r_2} \ldots x_k^{r_k}.$$

**(a)** How many terms are there in the sum?

**(b)** The sum of these multinomial coefficients has an easily expressed value:

$$\sum_{\substack{r_1+r_2+\cdots+r_k=n,\\ r_i\in\mathbb{N}}} \binom{n}{r_1,\ r_2,\ \ldots,\ r_k} = k^n \tag{15.24}$$

Give a combinatorial proof of this identity.

*Hint:* How many terms are there when $(x_1 + x_2 + \cdots + x_k)^n$ is expressed as a sum of monomials in $x_i$ *before* terms with like powers of these variables are collected together under a single coefficient?

**Problem 15.59.**
You want to choose a team of $m$ people for your startup company from a pool of $n$ applicants, and from these $m$ people you want to choose $k$ to be the team managers. You took a Math for Computer Science subject, so you know you can do this in

$$\binom{n}{m}\binom{m}{k}$$

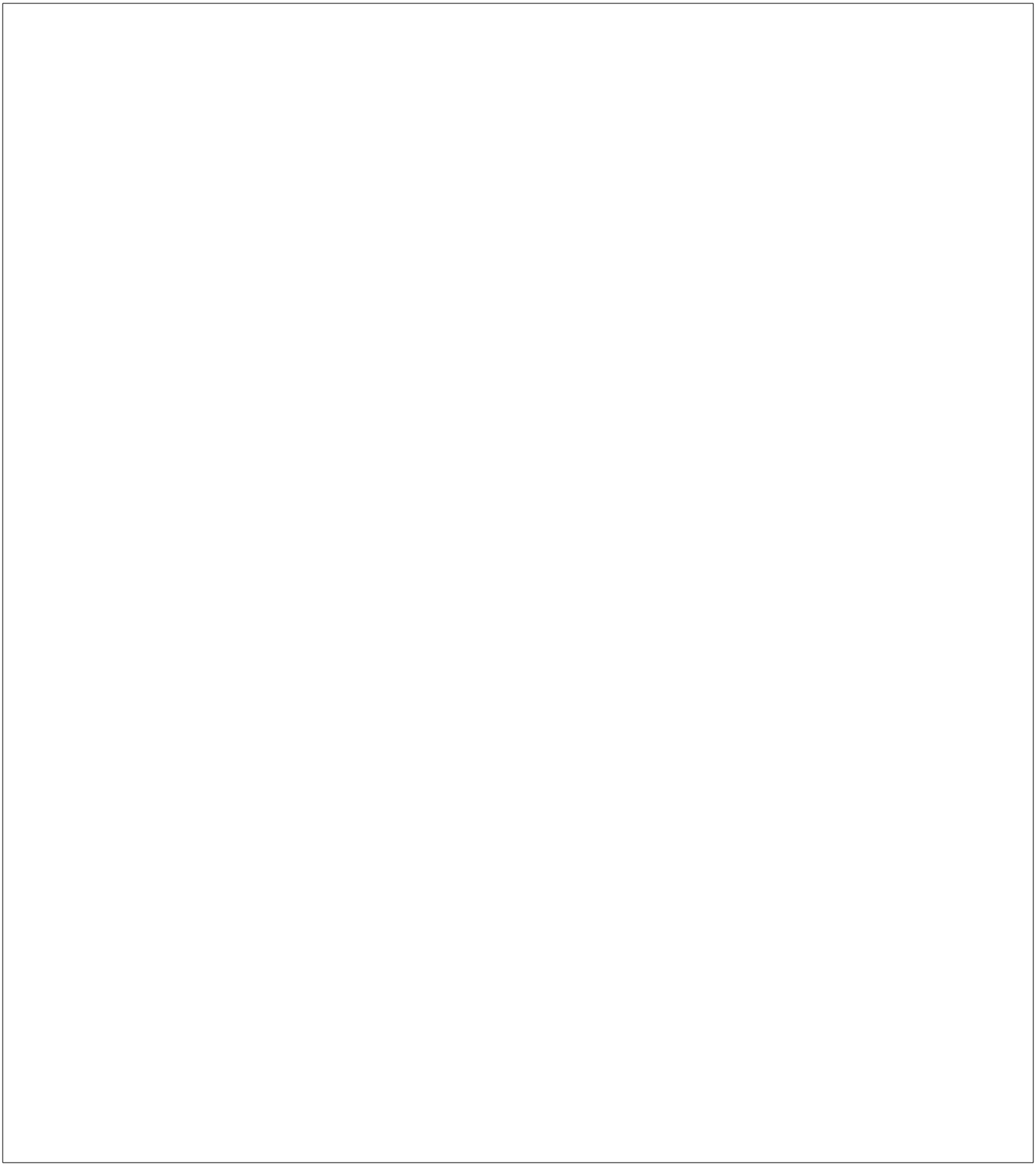
ways. But your CFO, who went to Harvard Business School, comes up with the formula

$$\binom{n}{k}\binom{n-k}{m-k}.$$

Before doing the reasonable thing—dump on your CFO or Harvard Business School—you decide to check his answer against yours.

**(a)** Give a *combinatorial proof* that your CFO's formula agrees with yours.

**(b)** Verify this combinatorial proof by giving an *algebraic* proof of this same fact.



# 16 Generating Functions

Generating Functions are one of the most surprising and useful inventions in Discrete Math. Roughly speaking, generating functions transform problems about *sequences* into problems about *functions*. This is great because we've got piles of mathematical machinery for manipulating functions. Thanks to generating functions, we can apply all that machinery to problems about sequences. In this way, we can use generating functions to solve all sorts of counting problems. There is a huge chunk of mathematics concerning generating functions, so we will only get a taste of the subject.

In this chapter, we'll put sequences in angle brackets to more clearly distinguish them from the many other mathematical expressions floating around.

The *ordinary generating function* for $\langle g_0, g_1, g_2, g_3 \dots \rangle$ is the power series:

$$G(x) = g_0 + g_1 x + g_2 x^2 + g_3 x^3 + \cdots .$$

There are a few other kinds of generating functions in common use, but ordinary generating functions are enough to illustrate the power of the idea, so we'll stick to them. So from now on *generating function* will mean the ordinary kind.

A generating function is a "formal" power series in the sense that we usually regard $x$ as a placeholder rather than a number. Only in rare cases will we actually evaluate a generating function by letting $x$ take a real number value, so we generally ignore the issue of convergence.

Throughout this chapter, we'll indicate the correspondence between a sequence and its generating function with a double-sided arrow as follows:

$$\langle g_0, g_1, g_2, g_3, \dots \rangle \longleftrightarrow g_0 + g_1 x + g_2 x^2 + g_3 x^3 + \cdots$$

For example, here are some sequences and their generating functions:

$$\begin{aligned}
\langle 0, 0, 0, 0, \dots \rangle &\longleftrightarrow 0 + 0x + 0x^2 + 0x^3 + \cdots = 0 \\
\langle 1, 0, 0, 0, \dots \rangle &\longleftrightarrow 1 + 0x + 0x^2 + 0x^3 + \cdots = 1 \\
\langle 3, 2, 1, 0, \dots \rangle &\longleftrightarrow 3 + 2x + 1x^2 + 0x^3 + \cdots = 3 + 2x + x^2
\end{aligned}$$

The pattern here is simple: the $i$th term in the sequence (indexing from 0) is the coefficient of $x^i$ in the generating function.

Recall that the sum of an infinite geometric series is:

$$1 + z + z^2 + z^3 + \cdots = \frac{1}{1 - z}$$

This equation does not hold when $|z| \geq 1$, but as remarked, we don't worry about convergence issues. This formula gives closed form generating functions for a whole range of sequences. For example:

$$\langle 1,1,1,1,\ldots\rangle \quad \longleftrightarrow \quad 1+x+x^2+x^3+\cdots \qquad = \frac{1}{1-x}$$

$$\langle 1,-1,1,-1,\ldots\rangle \quad \longleftrightarrow \quad 1-x+x^2-x^3+x^4-\cdots \qquad = \frac{1}{1+x}$$

$$\left\langle 1,a,a^2,a^3,\ldots\right\rangle \quad \longleftrightarrow \quad 1+ax+a^2x^2+a^3x^3+\cdots \qquad = \frac{1}{1-ax}$$

$$\langle 1,0,1,0,1,0,\ldots\rangle \quad \longleftrightarrow \quad 1+x^2+x^4+x^6+\cdots \qquad = \frac{1}{1-x^2}$$

## 16.1 Operations on Generating Functions

The magic of generating functions is that we can carry out all sorts of manipulations on sequences by performing mathematical operations on their associated generating functions. Let's experiment with various operations and characterize their effects in terms of sequences.

### 16.1.1 Scaling

Multiplying a generating function by a constant scales every term in the associated sequence by the same constant. For example, we noted above that:

$$\langle 1,0,1,0,1,0,\ldots\rangle \longleftrightarrow 1+x^2+x^4+x^6+\cdots = \frac{1}{1-x^2}$$

Multiplying the generating function by 2 gives

$$\frac{2}{1-x^2} = 2+2x^2+2x^4+2x^6+\cdots$$

which generates the sequence:

$$\langle 2,0,2,0,2,0,\ldots\rangle$$

**Rule 1** (Scaling Rule)**.** *If*

$$\langle f_0, f_1, f_2, \ldots\rangle \longleftrightarrow F(x),$$

*then*

$$\langle cf_0,\ cf_1,\ cf_2,\ \ldots\rangle \longleftrightarrow c\cdot F(x).$$

The idea behind this rule is that:

$$\begin{aligned}\langle cf_0, cf_1, cf_2, \ldots\rangle &\longleftrightarrow cf_0 + cf_1x + cf_2x^2 + \cdots\\ &= c\cdot(f_0 + f_1x + f_2x^2 + \cdots)\\ &= cF(x)\end{aligned}$$

### 16.1.2 Addition

Adding generating functions corresponds to adding the two sequences term by term. For example, adding two of our earlier examples gives:

$$\begin{array}{rccccccccc}
 & \langle & 1, & 1, & 1, & 1, & 1, & 1, & \ldots & \rangle \\
+ & \langle & 1, & -1, & 1, & -1, & 1, & -1, & \ldots & \rangle \\
\hline
 & \langle & 2, & 0, & 2, & 0, & 2, & 0, & \ldots & \rangle
\end{array}
\begin{array}{c}\longleftrightarrow\\ \longleftrightarrow\\ \hline \longleftrightarrow\end{array}
\begin{array}{c}\dfrac{1}{1-x}\\ \dfrac{1}{1+x}\\ \hline \dfrac{1}{1-x}+\dfrac{1}{1+x}\end{array}$$

We've now derived two different expressions that both generate the sequence $\langle 2, 0, 2, 0, \ldots\rangle$. They are, of course, equal:

$$\frac{1}{1-x} + \frac{1}{1+x} = \frac{(1+x)+(1-x)}{(1-x)(1+x)} = \frac{2}{1-x^2}$$

**Rule 2** (Addition Rule)**.** *If*

$$\begin{aligned}\langle f_0, f_1, f_2, \ldots\rangle &\longleftrightarrow F(x), \qquad \textit{and}\\ \langle g_0, g_1, g_2, \ldots\rangle &\longleftrightarrow G(x),\end{aligned}$$

*then*

$$\langle f_0 + g_0,\ f_1 + g_1,\ f_2 + g_2,\ \ldots\rangle \longleftrightarrow F(x) + G(x).$$

The idea behind this rule is that:

$$\begin{aligned}\langle f_0 + g_0,\ f_1 + g_1,\ f_2 + g_2,\ \ldots\rangle &\longleftrightarrow \sum_{n=0}^{\infty}(f_n + g_n)x^n\\ &= \left(\sum_{n=0}^{\infty} f_nx^n\right) + \left(\sum_{n=0}^{\infty} g_nx^n\right)\\ &= F(x) + G(x)\end{aligned}$$

### 16.1.3 Right Shifting

Let's start over again with a simple sequence and its generating function:

$$\langle 1, 1, 1, 1, \ldots \rangle \longleftrightarrow \frac{1}{1-x}$$

Now let's *right-shift* the sequence by adding $k$ leading zeros:

$$\begin{aligned} \langle \underbrace{0, 0, \ldots, 0}_{k \text{ zeroes}}, 1, 1, 1, \ldots \rangle &\longleftrightarrow x^k + x^{k+1} + x^{k+2} + x^{k+3} + \cdots \\ &= x^k \cdot (1 + x + x^2 + x^3 + \cdots) \\ &= \frac{x^k}{1-x} \end{aligned}$$

Evidently, adding $k$ leading zeros to the sequence corresponds to multiplying the generating function by $x^k$. This holds true in general.

**Rule 3** (Right-Shift Rule)**.** *If* $\langle f_0, f_1, f_2, \ldots \rangle \longleftrightarrow F(x)$*, then:*

$$\langle \underbrace{0, 0, \ldots, 0}_{k \text{ zeroes}}, f_0, f_1, f_2, \ldots \rangle \longleftrightarrow x^k \cdot F(x)$$

The idea behind this rule is that:

$$\begin{aligned} \langle \overbrace{0, 0, \ldots, 0}^{k \text{ zeroes}}, f_0, f_1, f_2, \ldots \rangle &\longleftrightarrow f_0 x^k + f_1 x^{k+1} + f_2 x^{k+2} + \cdots \\ &= x^k \cdot (f_0 + f_1 x + f_2 x^2 + f_3 x^3 + \cdots) \\ &= x^k \cdot F(x) \end{aligned}$$

### 16.1.4 Differentiation

What happens if we take the *derivative* of a generating function? As an example, let's differentiate the now-familiar generating function for an infinite sequence of 1's.

$$\begin{aligned} \frac{d}{dx}(1 + x + x^2 + x^3 + x^4 + \cdots) &= \frac{d}{dx}\left(\frac{1}{1-x}\right) \\ 1 + 2x + 3x^2 + 4x^3 + \cdots &= \frac{1}{(1-x)^2} \qquad (16.1) \\ \langle 1, 2, 3, 4, \ldots \rangle &\longleftrightarrow \frac{1}{(1-x)^2} \end{aligned}$$

We found a generating function for the sequence $\langle 1, 2, 3, 4, \ldots \rangle$ of positive integers!

In general, differentiating a generating function has two effects on the corresponding sequence: each term is multiplied by its index and the entire sequence is shifted left one place.

**Rule 4** (Derivative Rule)**.** *If*

$$\langle f_0, f_1, f_2, f_3, \ldots \rangle \longleftrightarrow F(x),$$

*then*

$$\langle f_1, 2f_2, 3f_3, \ldots \rangle \longleftrightarrow F'(x).$$

The idea behind this rule is that:

$$\begin{aligned} \langle f_1, 2f_2, 3f_3, \ldots \rangle &\longleftrightarrow f_1 + 2f_2 x + 3f_3 x^2 + \cdots \\ &= \frac{d}{dx}\left(f_0 + f_1 x + f_2 x^2 + f_3 x^3 + \cdots\right) \\ &= \frac{d}{dx} F(x) \end{aligned}$$

The Derivative Rule is very useful. In fact, there is frequent, independent need for each of differentiation's two effects, multiplying terms by their index and left-shifting one place. Typically, we want just one effect and must somehow cancel out the other. For example, let's try to find the generating function for the sequence of squares, $\langle 0, 1, 4, 9, 16, \ldots \rangle$. If we could start with the sequence $\langle 1, 1, 1, 1, \ldots \rangle$ and multiply each term by its index two times, then we'd have the desired result:

$$\langle 0 \cdot 0,\ 1 \cdot 1,\ 2 \cdot 2,\ 3 \cdot 3,\ \ldots \rangle = \langle 0, 1, 4, 9, \ldots \rangle$$

A challenge is that differentiation not only multiplies each term by its index, but also shifts the whole sequence left one place. However, the Right-Shift Rule 3 tells how to cancel out this unwanted left-shift: multiply the generating function by $x$.

Our procedure, therefore, is to begin with the generating function for $\langle 1, 1, 1, 1, \ldots \rangle$,

differentiate, multiply by $x$, and then differentiate and multiply by $x$ once more.

$$\begin{aligned}
\langle 1,1,1,1,\ldots\rangle &\longleftrightarrow \frac{1}{1-x}\\
\langle 1,2,3,4,\ldots\rangle &\longleftrightarrow \frac{d}{dx}\,\frac{1}{1-x} = \frac{1}{(1-x)^2}\\
\langle 0,1,2,3,\ldots\rangle &\longleftrightarrow x\cdot\frac{1}{(1-x)^2} = \frac{x}{(1-x)^2}\\
\langle 1,4,9,16,\ldots\rangle &\longleftrightarrow \frac{d}{dx}\,\frac{x}{(1-x)^2} = \frac{1+x}{(1-x)^3}\\
\langle 0,1,4,9,\ldots\rangle &\longleftrightarrow x\cdot\frac{1+x}{(1-x)^3} = \frac{x(1+x)}{(1-x)^3}
\end{aligned}$$

Thus, the generating function for squares is:

$$\frac{x(1+x)}{(1-x)^3} \tag{16.2}$$

### 16.1.5 Products

**Rule 5** ( Product Rule)**.** *If*

$$\langle a_0,a_1,a_2,\ldots\rangle \longleftrightarrow A(x), \quad \textit{and} \quad \langle b_0,b_1,b_2,\ldots\rangle \longleftrightarrow B(x),$$

*then*

$$\langle c_0,\ c_1,\ c_2,\ \ldots\rangle \longleftrightarrow A(x)\cdot B(x),$$

where

$$c_n ::= a_0b_n + a_1b_{n-1} + a_2b_{n-2} + \cdots + a_nb_0.$$

To understand this rule, let

$$C(x) ::= A(x)\cdot B(x) = \sum_{n=0}^{\infty} c_n x^n.$$

We can evaluate the product $A(x)\cdot B(x)$ by using a table to identify all the

cross-terms from the product of the sums:

| | $b_0x^0$ | $b_1x^1$ | $b_2x^2$ | $b_3x^3$ | ... |
|---|---|---|---|---|---|
| $a_0x^0$ | $a_0b_0x^0$ | $a_0b_1x^1$ | $a_0b_2x^2$ | $a_0b_3x^3$ | ... |
| $a_1x^1$ | $a_1b_0x^1$ | $a_1b_1x^2$ | $a_1b_2x^3$ | ... | |
| $a_2x^2$ | $a_2b_0x^2$ | $a_2b_1x^3$ | ... | | |
| $a_3x^3$ | $a_3b_0x^3$ | ... | | | |
| $\vdots$ | ... | | | | |

Notice that all terms involving the same power of $x$ lie on a /-sloped diagonal. Collecting these terms together, we find that the coefficient of $x^n$ in the product is the sum of all the terms on the $(n+1)$st diagonal, namely,

$$a_0b_n + a_1b_{n-1} + a_2b_{n-2} + \cdots + a_nb_0. \tag{16.3}$$

This expression (16.3) may be familiar from a signal processing course; the sequence $\langle c_0, c_1, c_2, \ldots \rangle$ is called the *convolution* of sequences $\langle a_0, a_1, a_2, \ldots \rangle$ and $\langle b_0, b_1, b_2, \ldots \rangle$.

## 16.2 The Fibonacci Sequence

Sometimes we can find nice generating functions for more complicated sequences. For example, here is a generating function for the Fibonacci numbers:

$$\langle 0, 1, 1, 2, 3, 5, 8, 13, 21, \ldots \rangle \longleftrightarrow \frac{x}{1 - x - x^2}$$

The Fibonacci numbers may seem like a fairly nasty bunch, but the generating function is simple!

We're going to derive this generating function and then use it to find a closed form for the $n$th Fibonacci number. The techniques we'll use are applicable to a large class of recurrence equations.

### 16.2.1 Finding a Generating Function

Let's begin by recalling the definition of the Fibonacci numbers:

$$\begin{aligned} f_0 &= 0 \\ f_1 &= 1 \\ f_n &= f_{n-1} + f_{n-2} \qquad (\text{for } n \geq 2) \end{aligned}$$

We can expand the final clause into an infinite sequence of equations. Thus, the Fibonacci numbers are defined by:

$$\begin{aligned} f_0 &= 0 \\ f_1 &= 1 \\ f_2 &= f_1 + f_0 \\ f_3 &= f_2 + f_1 \\ f_4 &= f_3 + f_2 \\ &\vdots \end{aligned}$$

Now the overall plan is to *define* a function $F(x)$ that generates the sequence on the left side of the equality symbols, which are the Fibonacci numbers. Then we *derive* a function that generates the sequence on the right side. Finally, we equate the two and solve for $F(x)$. Let's try this. First, we define:

$$F(x) = f_0 + f_1 x + f_2 x^2 + f_3 x^3 + f_4 x^4 + \cdots$$

Now we need to derive a generating function for the sequence:

$$\langle 0,\ 1,\ f_1 + f_0,\ f_2 + f_1,\ f_3 + f_2,\ \ldots \rangle$$

One approach is to break this into a sum of three sequences for which we know generating functions and then apply the Addition Rule:

$$\begin{array}{cccccccccc} & \langle & 0, & 1, & 0, & 0, & 0, & \ldots & \rangle & \longleftrightarrow & x \\ & \langle & 0, & f_0, & f_1, & f_2, & f_3, & \ldots & \rangle & \longleftrightarrow & xF(x) \\ + & \langle & 0, & 0, & f_0, & f_1, & f_2, & \ldots & \rangle & \longleftrightarrow & x^2F(x) \\ \hline & \langle & 0, & 1 + f_0, & f_1 + f_0, & f_2 + f_1, & f_3 + f_2, & \ldots & \rangle & \longleftrightarrow & x + xF(x) + x^2F(x) \end{array}$$

This sequence is almost identical to the right sides of the Fibonacci equations. The one blemish is that the second term is $1 + f_0$ instead of simply 1. However, this amounts to nothing, since $f_0 = 0$ anyway.

Now if we equate $F(x)$ with the new function $x + xF(x) + x^2F(x)$, then we're implicitly writing down *all* of the equations that define the Fibonacci numbers in one fell swoop:

$$\begin{array}{rcl} F(x) & = & f_0 + f_1 x + f_2 x^2 + f_3 x^3 + \cdots \\ \| & & \| \quad \| \quad \| \quad \| \\ x + xF(x) + x^2F(x) & = & 0 + (1 + f_0)\, x + (f_1 + f_0)\, x^2 + (f_2 + f_1)\, x^3 + \cdots \end{array}$$

Solving for $F(x)$ gives the generating function for the Fibonacci sequence:

$$F(x) = x + xF(x) + x^2F(x)$$

so

$$F(x) = \frac{x}{1 - x - x^2}.$$

Sure enough, this is the simple generating function we claimed at the outset.

### 16.2.2 Finding a Closed Form

Why should one care about the generating function for a sequence? There are several answers, but here is one: if we can find a generating function for a sequence, then we can often find a closed form for the $n$th coefficient —which can be pretty useful! For example, a closed form for the coefficient of $x^n$ in the power series for $x/(1 - x - x^2)$ would be an explicit formula for the $n$th Fibonacci number.

So our next task is to extract coefficients from a generating function. There are several approaches. For a generating function that is a ratio of polynomials, we can use the method of *partial fractions*, which you learned in calculus. Just as the terms in a partial fraction expansion are easier to integrate, the coefficients of those terms are easy to compute.

Let's try this approach with the generating function for Fibonacci numbers. First, we factor the denominator:

$$1 - x - x^2 = (1 - \alpha_1 x)(1 - \alpha_2 x)$$

where $\alpha_1 = \frac{1}{2}(1 + \sqrt{5})$ and $\alpha_2 = \frac{1}{2}(1 - \sqrt{5})$. Next, we find $A_1$ and $A_2$ which satisfy:

$$\frac{x}{1 - x - x^2} = \frac{A_1}{1 - \alpha_1 x} + \frac{A_2}{1 - \alpha_2 x}$$

We do this by plugging in various values of $x$ to generate linear equations in $A_1$ and $A_2$. We can then find $A_1$ and $A_2$ by solving a linear system. This gives:

$$A_1 = \frac{1}{\alpha_1 - \alpha_2} = \frac{1}{\sqrt{5}}$$

$$A_2 = \frac{-1}{\alpha_1 - \alpha_2} = -\frac{1}{\sqrt{5}}$$

Substituting into the equation above gives the partial fractions expansion of $F(x)$:

$$\frac{x}{1-x-x^2} = \frac{1}{\sqrt{5}}\left(\frac{1}{1-\alpha_1 x} - \frac{1}{1-\alpha_2 x}\right)$$

Each term in the partial fractions expansion has a simple power series given by the geometric sum formula:

$$\frac{1}{1-\alpha_1 x} = 1 + \alpha_1 x + \alpha_1^2 x^2 + \cdots$$
$$\frac{1}{1-\alpha_2 x} = 1 + \alpha_2 x + \alpha_2^2 x^2 + \cdots$$

Substituting in these series gives a power series for the generating function:

$$\begin{aligned} F(x) &= \frac{1}{\sqrt{5}}\left(\frac{1}{1-\alpha_1 x} - \frac{1}{1-\alpha_2 x}\right) \\ &= \frac{1}{\sqrt{5}}\left((1 + \alpha_1 x + \alpha_1^2 x^2 + \cdots) - (1 + \alpha_2 x + \alpha_2^2 x^2 + \cdots)\right), \end{aligned}$$

so

$$\begin{aligned} f_n &= \frac{\alpha_1^n - \alpha_2^n}{\sqrt{5}} \\ &= \frac{1}{\sqrt{5}}\left(\left(\frac{1+\sqrt{5}}{2}\right)^n - \left(\frac{1-\sqrt{5}}{2}\right)^n\right) \end{aligned}$$

This formula may be scary and astonishing —it's not even obvious that its value is an integer —but it's very useful. For example, it provides (via the repeated squaring method) a much more efficient way to compute Fibonacci numbers than crunching through the recurrence, and it also clearly reveals the exponential growth of these numbers.

## 16.3 Counting with Generating Functions

Generating functions are particularly useful for solving counting problems. In particular, problems involving choosing items from a set often lead to nice generating functions by letting the coefficient of $x^n$ be the number of ways to choose $n$ items.

### 16.3.1 Choosing Distinct Items from a Set

The generating function for binomial coefficients follows directly from the Binomial Theorem:

$$\left\langle \binom{k}{0}, \binom{k}{1}, \binom{k}{2}, \ldots, \binom{k}{k}, 0, 0, 0, \ldots \right\rangle \quad \longleftrightarrow \quad \binom{k}{0} + \binom{k}{1}x + \binom{k}{2}x^2 + \cdots + \binom{k}{k}x^k$$
$$= \quad (1+x)^k$$

Thus, the coefficient of $x^n$ in $(1+x)^k$ is $\binom{k}{n}$, the number of ways to choose $n$ distinct items from a set of size $k$. For example, the coefficient of $x^2$ is $\binom{k}{2}$, the number of ways to choose 2 items from a set with $k$ elements. Similarly, the coefficient of $x^{k+1}$ is the number of ways to choose $k+1$ items from a size $k$ set, which is zero. (Watch out for this reversal of the roles that $k$ and $n$ played in earlier examples; we're led to this reversal because we've been using $n$ to refer to the power of $x$ in a power series.)

### 16.3.2 Building Generating Functions that Count

Often we can translate the description of a counting problem directly into a generating function for the solution. For example, we could figure out that $(1+x)^k$ generates the number of ways to select $n$ distinct items from a $k$-element set without resorting to the Binomial Theorem or even fussing with binomial coefficients!

Here is how. First, consider a single-element set $\{a_1\}$. The generating function for the number of ways to select $n$ elements from this set is simply $1+x$: we have 1 way to select zero elements, 1 way to select one element, and 0 ways to select more than one element. Similarly, the number of ways to select $n$ elements from the set $\{a_2\}$ is also given by the generating function $1+x$. The fact that the elements differ in the two cases is irrelevant.

Now here is the main trick: *the generating function for choosing elements from a union of disjoint sets is the product of the generating functions for choosing from each set.* We'll justify this in a moment, but let's first look at an example. According to this principle, the generating function for the number of ways to select $n$ elements from the $\{a_1, a_2\}$ is:

$$\underbrace{(1+x)}_{\substack{\text{gen func for}\\ \text{selecting an } a_1}} \cdot \underbrace{(1+x)}_{\substack{\text{gen func for}\\ \text{selecting an } a_2}} = \underbrace{(1+x)^2}_{\substack{\text{gen func for}\\ \text{selecting from}\\ \{a_1, a_2\}}} = 1 + 2x + x^2$$

Sure enough, for the set $\{a_1, a_2\}$, we have 1 way to select zero elements, 2 ways to

select one element, 1 way to select two elements, and 0 ways to select more than two elements.

Repeated application of this rule gives the generating function for selecting $n$ items from a $k$-element set $\{a_1, a_2, \ldots, a_k\}$:

$$\underbrace{(1+x)}_{\substack{\text{gen func for}\\ \text{selecting an } a_1}} \cdot \underbrace{(1+x)}_{\substack{\text{gen func for}\\ \text{selecting an } a_2}} \cdots \underbrace{(1+x)}_{\substack{\text{gen func for}\\ \text{selecting an } a_k}} = \underbrace{(1+x)^k}_{\substack{\text{gen func for}\\ \text{selecting from}\\ \{a_1, a_2, \ldots, a_k\}}}$$

This is the same generating function that we obtained by using the Binomial Theorem. But this time around we translated directly from the counting problem to the generating function.

We can extend these ideas to a general principle:

**Rule 6** (Convolution Rule)**.** *Let $A(x)$ be the generating function for selecting items from set $\mathcal{A}$, and let $B(x)$ be the generating function for selecting items from set $\mathcal{B}$. If $\mathcal{A}$ and $\mathcal{B}$ are disjoint, then the generating function for selecting items from the union $\mathcal{A} \cup \mathcal{B}$ is the product $A(x) \cdot B(x)$.*

This rule is rather ambiguous: what exactly are the rules governing the selection of items from a set? Remarkably, the Convolution Rule remains valid under *many* interpretations of selection. For example, we could insist that distinct items be selected or we might allow the same item to be picked a limited number of times or any number of times. Informally, the only restrictions are that (1) the order in which items are selected is disregarded and (2) restrictions on the selection of items from sets $\mathcal{A}$ and $\mathcal{B}$ also apply in selecting items from $\mathcal{A} \cup \mathcal{B}$. (Formally, there must be a bijection between $n$-element selections from $\mathcal{A} \cup \mathcal{B}$ and ordered pairs of selections from $\mathcal{A}$ and $\mathcal{B}$ containing a total of $n$ elements.)

To count the number of ways to select $n$ items from $\mathcal{A} \cup \mathcal{B}$, we observe that we can select $n$ items by choosing $j$ items from $\mathcal{A}$ and $n-j$ items from $\mathcal{B}$, where $j$ is any number from 0 to $n$. This can be done in $a_j b_{n-j}$ ways. Summing over all the possible values of $j$ gives a total of

$$a_0 b_n + a_1 b_{n-1} + a_2 b_{n-2} + \cdots + a_n b_0$$

ways to select $n$ items from $\mathcal{A} \cup \mathcal{B}$. By the Product Rule, this is precisely the coefficient of $x^n$ in the series for $A(x)B(x)$.

### 16.3.3 Choosing Items with Repetition

The first counting problem we considered was the number of ways to select a dozen doughnuts when five flavors were available. We can generalize this question as

follows: in how many ways can we select $n$ items from a $k$-element set if we're allowed to pick the same item multiple times? In these terms, the doughnut problem asks in how many ways we can select $n = 12$ doughnuts from the set of $k = 5$ flavors

$$\{\text{chocolate, lemon-filled, sugar, glazed, plain}\}$$

where, of course, we're allowed to pick several doughnuts of the same flavor. Let's approach this question from a generating functions perspective.

Suppose we make $n$ choices (with repetition allowed) of items from a set containing a single item. Then there is one way to choose zero items, one way to choose one item, one way to choose two items, etc. Thus, the generating function for choosing $n$ elements with repetition from a 1-element set is:

$$\begin{aligned}\langle 1,1,1,1,\ldots\rangle &\longleftrightarrow 1 + x + x^2 + x^3 + \cdots \\ &= \frac{1}{1-x}\end{aligned}$$

The Convolution Rule says that the generating function for selecting items from a union of disjoint sets is the product of the generating functions for selecting items from each set:

$$\underbrace{\frac{1}{1-x}}_{\substack{\text{gen func for}\\ \text{choosing } a_1\text{'s}}} \cdot \underbrace{\frac{1}{1-x}}_{\substack{\text{gen func for}\\ \text{choosing } a_2\text{'s}}} \cdots \underbrace{\frac{1}{1-x}}_{\substack{\text{gen func for}\\ \text{choosing } a_k\text{'s}}} = \underbrace{\frac{1}{(1-x)^k}}_{\substack{\text{gen func for}\\ \text{repeated choice from}\\ \{a_1, a_2, \ldots, a_k\}}}$$

Therefore, the generating function for choosing items from a $k$-element set with repetition allowed is $1/(1-x)^k$.

Now the Bookkeeper Rule tells us that the number of ways to choose $n$ items with repetition from an $k$ element set is

$$\binom{n+k-1}{n},$$

so this is the coefficient of $x^n$ in the series expansion of $1/(1-x)^k$.

On the other hand, it's instructive to derive this coefficient algebraically, which we can do using Taylor's Theorem:

**Theorem 16.3.1** (Taylor's Theorem)**.**

$$f(x) = f(0) + f'(0)x + \frac{f''(0)}{2!}x^2 + \frac{f'''(0)}{3!}x^3 + \cdots + \frac{f^{(n)}(0)}{n!}x^n + \cdots.$$

This theorem says that the $n$th coefficient of $1/(1-x)^k$ is equal to its $n$th derivative evaluated at 0 and divided by $n!$. Computing the $n$th derivative turns out not to be very difficult (Problem 16.9).

## 16.4 An "Impossible" Counting Problem

So far everything we've done with generating functions we could have done another way. But here is an absurd counting problem —really over the top! In how many ways can we fill a bag with $n$ fruits subject to the following constraints?

- The number of apples must be even.
- The number of bananas must be a multiple of 5.
- There can be at most four oranges.
- There can be at most one pear.

For example, there are 7 ways to form a bag with 6 fruits:

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Apples | 6 | 4 | 4 | 2 | 2 | 0 | 0 |
| Bananas | 0 | 0 | 0 | 0 | 0 | 5 | 5 |
| Oranges | 0 | 2 | 1 | 4 | 3 | 1 | 0 |
| Pears | 0 | 0 | 1 | 0 | 1 | 0 | 1 |

These constraints are so complicated that the problem may seem hopeless. But let's see what generating functions reveal.

Let's first construct a generating function for choosing apples. We can choose a set of 0 apples in one way, a set of 1 apple in zero ways (since the number of apples must be even), a set of 2 apples in one way, a set of 3 apples in zero ways, and so forth. So we have:

$$A(x) = 1 + x^2 + x^4 + x^6 + \cdots = \frac{1}{1-x^2}$$

Similarly, the generating function for choosing bananas is:

$$B(x) = 1 + x^5 + x^{10} + x^{15} + \cdots = \frac{1}{1-x^5}$$

Now, we can choose a set of 0 oranges in one way, a set of 1 orange in one way, and so on. However, we can not choose more than four oranges, so we have the

generating function:

$$O(x) = 1 + x + x^2 + x^3 + x^4 = \frac{1 - x^5}{1 - x}$$

Here we're using the geometric sum formula. Finally, we can choose only zero or one pear, so we have:

$$P(x) = 1 + x$$

The Convolution Rule says that the generating function for choosing from among all four kinds of fruit is:

$$\begin{aligned} A(x)B(x)O(x)P(x) &= \frac{1}{1-x^2}\frac{1}{1-x^5}\frac{1-x^5}{1-x}(1+x) \\ &= \frac{1}{(1-x)^2} \\ &= 1 + 2x + 3x^2 + 4x^3 + \cdots \end{aligned}$$

Almost everything cancels! We're left with $1/(1-x)^2$, which we found a power series for earlier: the coefficient of $x^n$ is simply $n+1$. Thus, the number of ways to form a bag of $n$ fruits is just $n + 1$. This is consistent with the example we worked out, since there were 7 different fruit bags containing 6 fruits. *Amazing!*

## Problems for Section 16.2

### Class Problems

**Problem 16.1.**

The famous mathematician, Fibonacci, has decided to start a rabbit farm to fill up his time while he's not making new sequences to torment future college students. Fibonacci starts his farm on month zero (being a mathematician), and at the start of month one he receives his first pair of rabbits. Each pair of rabbits takes a month to mature, and after that breeds to produce one new pair of rabbits each month. Fibonacci decides that in order never to run out of rabbits or money, every time a batch of new rabbits is born, he'll sell a number of newborn pairs equal to the total number of pairs he had three months earlier. Fibonacci is convinced that this way he'll never run out of stock.

**(a)** Define the number, $r_n$, of pairs of rabbits Fibonacci has in month $n$, using a recurrence relation. That is, define $r_n$ in terms of various $r_i$ where $i < n$.

**(b)** Let $R(x)$ be the generating function for rabbit pairs,

$$R(x) ::= r_0 + r_1 x + r_2 x^2 + \cdot.$$

Express $R(x)$ as a quotient of polynomials.

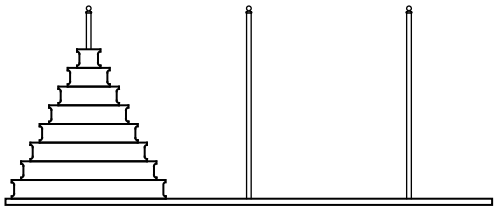


**Figure 16.1** The initial configuration of the disks in the Towers of Sheboygan problem.

**(c)** Find a partial fraction decomposition of the generating function $R(x)$.

**(d)** Finally, use the partial fraction decomposition to come up with a closed form expression for the number of pairs of rabbits Fibonacci has on his farm on month $n$.

**Problem 16.2.**

Less well-known than the Towers of Hanoi —but no less fascinating —are the Towers of Sheboygan. As in Hanoi, the puzzle in Sheboygan involves 3 posts and $n$ disks of different sizes. Initially, all the disks are on post #1 as in Figure 16.1.

The objective is to transfer all $n$ disks to post #2 via a sequence of moves. A move consists of removing the top disk from one post and dropping it onto another post with the restriction that a larger disk can never lie above a smaller disk. Furthermore, a local ordinance requires that *a disk can be moved only from a post to the next post on its right —or from post #3 to post #1.* Thus, for example, moving a disk directly from post #1 to post #3 is not permitted.

**(a)** One procedure that solves the Sheboygan puzzle is defined recursively: to move an initial stack of $n$ disks to the next post, move the top stack of $n-1$ disks to the furthest post by moving it to the next post two times, then move the big, $n$th disk to the next post, and finally move the top stack another two times to land on top of the big disk. Let $s_n$ be the number of moves that this procedure uses. Write a simple linear recurrence for $s_n$.

**(b)** Let $S(x)$ be the generating function for the sequence $\langle s_0, s_1, s_2, \ldots \rangle$. Carefully Show that

$$S(x) = \frac{x}{(1-x)(1-4x)}.$$

**(c)** Give a simple formula for $s_n$.

**(d)** A better (indeed optimal, but we won't prove this) procedure to solve the Towers of Sheboygan puzzle can be defined in terms of two mutually recursive procedures, procedure $P_1(n)$ for moving a stack of $n$ disks 1 pole forward, and $P_2(n)$ for moving a stack of $n$ disks 2 poles forward. This is trivial for $n = 0$. For $n > 0$, define:

$P_1(n)$: Apply $P_2(n-1)$ to move the top $n-1$ disks two poles forward to the third pole. Then move the remaining big disk once to land on the second pole. Then apply $P_2(n-1)$ again to move the stack of $n-1$ disks two poles forward from the third pole to land on top of the big disk.

$P_2(n)$: Apply $P_2(n-1)$ to move the top $n-1$ disks two poles forward to land on the third pole. Then move the remaining big disk to the second pole. Then apply $P_1(n-1)$ to move the stack of $n-1$ disks one pole forward to land on the first pole. Now move the big disk 1 pole forward again to land on the third pole. Finally, apply $P_2(n-1)$ again to move the stack of $n-1$ disks two poles forward to land on the big disk.

Let $t_n$ be the number of moves needed to solve the Sheboygan puzzle using procedure $P_1(n)$. Show that

$$t_n = 2t_{n-1} + 2t_{n-2} + 3, \tag{16.4}$$

for $n > 1$.

*Hint:* Let $u_n$ be the number of moves used by procedure $P_2(n)$. Express each of $t_n$ and $u_n$ as linear combinations of $t_{n-1}$ and $u_{n-1}$ and solve for $t_n$.

**(e)** Derive values $a, b, c, \alpha, \beta$ such that

$$t_n = a\alpha^n + b\beta^n + c.$$

Conclude that $t_n = o(s_n)$.

## Homework Problems

**Problem 16.3.**
Taking derivatives of generating functions is another useful operation. This is done termwise, that is, if

$$F(x) = f_0 + f_1 x + f_2 x^2 + f_3 x^3 + \cdots,$$

then

$$F'(x) ::= f_1 + 2f_2 x + 3f_3 x^2 + \cdots.$$

For example,

$$\frac{1}{(1-x)^2} = \left(\frac{1}{(1-x)}\right)' = 1 + 2x + 3x^2 + \cdots$$

so

$$H(x) ::= \frac{x}{(1-x)^2} = 0 + 1x + 2x^2 + 3x^3 + \cdots$$

is the generating function for the sequence of nonnegative integers. Therefore

$$\frac{1+x}{(1-x)^3} = H'(x) = 1 + 2^2x + 3^2x^2 + 4^2x^3 + \cdots,$$

so

$$\frac{x^2+x}{(1-x)^3} = xH'(x) = 0 + 1x + 2^2x^2 + 3^2x^3 + \cdots + n^2x^n + \cdots$$

is the generating function for the nonnegative integer squares.

**(a)** Prove that for all $k \in \mathbb{N}$, the generating function for the nonnegative integer $k$th powers is a quotient of polynomials in $x$. That is, for all $k \in \mathbb{N}$ there are polynomials $R_k(x)$ and $S_k(x)$ such that

$$[x^n]\left(\frac{R_k(x)}{S_k(x)}\right) = n^k. \tag{16.5}$$

*Hint:* Observe that the derivative of a quotient of polynomials is also a quotient of polynomials. It is not necessary work out explicit formulas for $R_k$ and $S_k$ to prove this part.

**(b)** Conclude that if $f(n)$ is a function on the nonnegative integers defined recursively in the form

$$f(n) = af(n-1) + bf(n-2) + cf(n-3) + p(n)\alpha^n$$

where the $a, b, c, \alpha \in \mathbb{C}$ and $p$ is a polynomial with complex coefficients, then the generating function for the sequence $f(0), f(1), f(2), \ldots$ will be a quotient of polynomials in $x$, and hence there is a closed form expression for $f(n)$.

*Hint:* Consider

$$\frac{R_k(\alpha x)}{S_k(\alpha x)}$$

**Problem 16.4.**
Generating functions provide an interesting way to count the number of strings of matched brackets. To do this, we'll use a description of these strings as the set, GoodCount, of strings of brackets with a good count.

Namely, one precise way to determine if a string is matched is to start with 0 and read the string from left to right, adding 1 to the count for each left bracket and subtracting 1 from the count for each right bracket. For example, here are the counts for the two strings above

| [ | ] | ] | [ | [ | [ | [ | [ | ] | ] | ] | ] |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 1 | 0 | −1 | 0 | 1 | 2 | 3 | 4 | 3 | 2 | 1 | 0 |

| [ | [ | [ | ] | ] | [ | ] | ] | [ | ] |
|---|---|---|---|---|---|---|---|---|---|
| 0 | 1 | 2 | 3 | 2 | 1 | 2 | 1 | 0 | 1 | 0 |

A string has a *good count* if its running count never goes negative and ends with 0. So the second string above has a good count, but the first one does not because its count went negative at the third step.

**Definition 16.4.1.** Let

$$\text{GoodCount} ::= \{s \in \{\texttt{]}, \texttt{[}\}^* \mid s \text{ has a good count}\}.$$

The matched strings can now be characterized precisely as this set of strings with good counts.

Let $c_n$ be the number of strings in GoodCount with exactly $n$ left brackets, and let $C(x)$ be the generating function for these numbers:

$$C(x) ::= c_0 + c_1 x + c_2 x^2 + \cdots.$$

**(a)** The *wrap* of a string, $s$, is the string, $\texttt{[}s\texttt{]}$, that starts with a left bracket followed by the characters of $s$, and then ends with a right bracket. Explain why the generating function for the wraps of strings with a good count is $xC(x)$.

*Hint:* The wrap of a string with good count also has a good count that starts and ends with 0 and remains *positive* everywhere else.

**(b)** Explain why, for every string, $s$, with a good count, there is a unique sequence of strings $s_1, \ldots, s_k$ that are wraps of strings with good counts and $s = s_1 \cdots s_k$. For example, the string $r ::= \texttt{[[]][][[]]} \in$ GoodCount equals $s_1 s_2 s_3$ where $s_1 ::= \texttt{[[]]}$, $s_2 ::= \texttt{[]}$, $s_3 ::= \texttt{[[]]}$, and this is the only way to express $r$ as a sequence of wraps of strings with good counts.

**(c)** Conclude that

$$C = 1 + xC + (xC)^2 + \cdots + (xC)^n + \cdots, \tag{16.6}$$

so

$$C = \frac{1}{1 - xC}, \tag{16.7}$$

and hence

$$C = \frac{1 \pm \sqrt{1 - 4x}}{2x}. \tag{16.8}$$

Let $D(x) ::= 2xC(x)$. Expressing $D$ as a power series

$$D(x) = d_0 + d_1 x + d_2 x^2 + \cdots,$$

we have

$$c_n = \frac{d_{n+1}}{2}. \tag{16.9}$$

**(d)** Use (16.8), (16.9), and the value of $c_0$ to conclude that

$$D(x) = 1 - \sqrt{1 - 4x}.$$

**(e)** Prove that

$$d_n = \frac{(2n-3)\cdot(2n-5)\cdots 5\cdot 3\cdot 1\cdot 2^n}{n!}.$$

*Hint:* $d_n = D^{(n)}(0)/n!$

**(f)** Conclude that

$$c_n = \frac{1}{n+1}\binom{2n}{n}.$$

## Exam Problems

**Problem 16.5.**
Define the sequence $r_0, r_1, r_2, \ldots$ recursively by the rule that $r_0 = r_1 = 0$ and

$$r_n = 7r_{n-1} + 4r_{n-2} + (n+1),$$

for $n \geq 2$. Express the generating function of this sequence as a quotient of polynomials or products of polynomials. You do *not* have to find a closed form for $r_n$.

## Problems for Section 16.3

### Practice Problems

**Problem 16.6.**
You would like to buy a bouquet of flowers. You find an online service that will make bouquets of **lilies**, **roses** and **tulips**, subject to the following constraints:

- there must be at most 3 lilies,
- there must be an odd number of tulips,
- there can be any number of roses.

Example: A bouquet of 3 tulips, 5 roses and no lilies satisfies the constraints.

Let $f_n$ be the number of possible bouquets with $n$ flowers that fit the service's constraints. Express $F(x)$, the generating function corresponding to $\langle f_0, f_1, f_2, \ldots \rangle$, as a quotient of polynomials (or products of polynomials). You do not need to simplify this expression.

**Problem 16.7.**
Let $b, c, a_0, a_1, a_2, \ldots$ be real numbers such that

$$a_n = b(a_{n-1}) + c$$

for $n \geq 1$.

Let $G(x)$ be the generating function for this sequence.

**(a)** The coefficient of $x^n$ in the series expansion of $G(x)$ is

**(b)** The coefficient of $x^n$ for $n \geq 1$ in the series expansion of $bxG(x)$ is

**(c)** The coefficient of $x^n$ for $n \geq 1$ in the series expansion of $cx/(1-x)$ is

**(d)** Therefore, $G(x) - bxG(x) - cx/(1-x) =$

**(e)** Using the method of partial fractions, we can find real numbers $d$ and $e$ such that

$$G(x) = d/L(x) + e/M(x).$$

What are $L(x)$ and $M(x)$?

**Problem 16.8.**
Write a formula for the generating function of each of the following sequences.

**(a)** 0, 0, 1, 1, 1,...

**(b)** 1, 1, 0, 0, 0,...

**(c)** 1, 0, 1, 0, 1, 0, 1,...

**(d)** 1, 4, 6, 4, 1, 0, 0, 0,...

**(e)** 1, 1, 1/2, 1/6, 1/24, 1/120,...

**(f)** 1, 2, 3, 4, 5,...

**(g)** 1, 4, 9, 16, 25,...

### Class Problems

**Problem 16.9.**
Let $A(x) = \sum_{n=0}^{\infty} a_n x^n$. Then it's easy to check that

$$a_n = \frac{A^{(n)}(0)}{n!},$$

where $A^{(n)}$ is the $n$th derivative of $A$. Use this fact (which you may assume) instead of the Convolution Counting Principle, to prove that

$$\frac{1}{(1-x)^k} = \sum_{n=0}^{\infty} \binom{n+k-1}{k-1} x^n.$$

So if we didn't already know the Bookkeeper Rule, we could have proved it from this calculation and the Convolution Rule for generating functions.

**Problem 16.10.**
We are interested in generating functions for the number of different ways to compose a bag of $n$ donuts subject to various restrictions. For each of the restrictions in (a)-(e) below, find a closed form for the corresponding generating function.

**(a)** All the donuts are chocolate and there are at least 3.

**(b)** All the donuts are glazed and there are at most 2.

**(c)** All the donuts are coconut and there are exactly 2 or there are none.

**(d)** All the donuts are plain and their number is a multiple of 4.

**(e)** The donuts must be chocolate, glazed, coconut, or plain and:

- there must be at least 3 chocolate donuts, and
- there must be at most 2 glazed, and
- there must be exactly 0 or 2 coconut, and
- there must be a multiple of 4 plain.

**(f)** Find a closed form for the number of ways to select $n$ donuts subject to the constraints of the previous part.

**Problem 16.11. (a)** Let

$$S(x) ::= \frac{x^2 + x}{(1-x)^3}.$$

What is the coefficient of $x^n$ in the generating function series for $S(x)$?

**(b)** Explain why $S(x)/(1-x)$ is the generating function for the sums of squares. That is, the coefficient of $x^n$ in the series for $S(x)/(1-x)$ is $\sum_{k=1}^{n} k^2$.

**(c)** Use the previous parts to prove that

$$\sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6}.$$

**Homework Problems**

**Problem 16.12.**
We will use generating functions to determine how many ways there are to use pennies, nickels, dimes, quarters, and half-dollars to give $n$ cents change.

**(a)** Write the sequence $P_n$ for the number of ways to use only pennies to change $n$ cents. Write the generating function for that sequence.

**(b)** Write the sequence $N_n$ for the number of ways to use only nickels to change $n$ cents. Write the generating function for that sequence.

**(c)** Write the generating function for the number of ways to use only nickels and pennies to change $n$ cents.

**(d)** Write the generating function for the number of ways to use pennies, nickels, dimes, quarters, and half-dollars to give $n$ cents change.

**(e)** Explain how to use this function to find out how many ways are there to change 50 cents; you do *not* have to provide the answer or actually carry out the process.

### Exam Problems

**Problem 16.13.**
Find the coefficients of $x^{10}y^5$ in $(19x + 4y)^{15}$

## Problems for Section 16.4

### Homework Problems

**Problem 16.14.**
Miss McGillicuddy never goes outside without a collection of pets. In particular:

- She brings a positive number of songbirds, which always come in pairs.
- She may or may not bring her alligator, Freddy.
- She brings at least 2 cats.
- She brings two or more chihuahuas and labradors leashed together in a line.

Let $P_n$ denote the number of different collections of $n$ pets that can accompany her, where we regard chihuahuas and labradors leashed up in different orders as different collections, even if there are the same number chihuahuas and labradors leashed in the line.

For example, $P_6 = 4$ since there are 4 possible collections of 6 pets:

- 2 songbirds, 2 cats, 2 chihuahuas leashed in line
- 2 songbirds, 2 cats, 2 labradors leashed in line
- 2 songbirds, 2 cats, a labrador leashed behind a chihuahua
- 2 songbirds, 2 cats, a chihuahua leashed behind a labrador

And $P_7 = 16$ since there are 16 possible collections of 7 pets:

- 2 songbirds, 3 cats, 2 chihuahuas leashed in line
- 2 songbirds, 3 cats, 2 labradors leashed in line
- 2 songbirds, 3 cats, a labrador leashed behind a chihuahua
- 2 songbirds, 3 cats, a chihuahua leashed behind a labrador

- 4 collections consisting of 2 songbirds, 2 cats, 1 alligator, and a line of 2 dogs
- 8 collections consisting of 2 songbirds, 2 cats, and a line of 3 dogs.

**(a)** Let

$$P(x) ::= P_0 + P_1 x + P_2 x^2 + P_3 x^3 + \cdots$$

be the generating function for the number of Miss McGillicuddy's pet collections. Verify that

$$P(x) = \frac{4x^6}{(1-x)^2(1-2x)}.$$

**(b)** Find a simple formula for $P_n$.

### Exam Problems

**Problem 16.15.**
T-Pain is planning an epic boat trip and he needs to decide what to bring with him.

- He must bring some burgers, but they only come in packs of 6.
- He and his two friends can't decide whether they want to dress formally or casually. He'll either bring 0 pairs of flip flops or 3 pairs.
- He doesn't have very much room in his suitcase for towels, so he can bring at most 2.
- In order for the boat trip to be truly epic, he has to bring at least 1 nautical-themed pashmina afghan.

**(a)** Let $B(x)$ be the generating function for the number of ways to bring $n$ burgers, $F(x)$ for the number of ways to bring $n$ pairs of flip flops, $T(x)$ for towels, and $A(x)$ for Afghans. Write simple formulas for each of these.

$$B(x): \qquad F(x):$$

$$T(x): \qquad A(x):$$

**(b)** Let $g_n$ be the the number of different ways for T-Pain to bring $n$ items (burgers, pairs of flip flops, towels, and/or afghans) on his boat trip. Let $G(x)$ be the generating function $\sum_{n=0}^{\infty} g_n x^n$. Verify that

$$G(x) = \frac{x^7}{(1-x)^2}.$$

**(c)** Find a simple formula for $g_n$.

# IV Probability

# Introduction

Probability is one of the most important disciplines in all of the sciences. It is also one of the least well understood.

Probability is especially important in computer science—it arises in virtually every branch of the field. In algorithm design and game theory, for example, *randomized* algorithms and strategies (those that use a random number generator as a key input for decision making) frequently outperform deterministic algorithms and strategies. In information theory and signal processing, an understanding of randomness is critical for filtering out noise and compressing data. In cryptography and digital rights management, probability is crucial for achieving security. The list of examples is long.

Given the impact that probability has on computer science, it seems strange that probability should be so misunderstood by so many. Perhaps the trouble is that basic human intuition is wrong as often as it is right when it comes to problems involving random events. As a consequence, many students develop a fear of probability. Indeed, we have witnessed many graduate oral exams where a student will solve the most horrendous calculation, only to then be tripped up by the simplest probability question. Indeed, even some faculty will start squirming if you ask them a question that starts "What is the probability that. . . ?"

Our goal in the remaining chapters is to equip you with the tools that will enable you to solve basic problems involving probability easily and confidently.

Chapter 17 introduces the basic definitions and an elementary 4-step process that can be used to determine the probability that a specified event occurs. We illustrate the method on two famous problems where your intuition will probably fail you. The key concepts of Conditional probability and independence are introduced, along with examples of their use, and regrettable misuse, in practice: the probability you have a disease given that a diagnostic test says you do, and the probability

that a suspect is guilty given that his blood type matches the blood found at the scene of the crime.

Random variables provide a more quantitative way to measure random events and We study them in Chapter 18. For example, instead of determining the probability that it will rain, we may want to determine *how much* or *how long* it is likely to rain. The fundamental concept of the *expected value* of a random variable is introduced and some of its key properties are developed.

Chapter 19 examines the probability that a random variable deviates significantly from its expected value. Probability of deviation provides the theoretical basis for estimation by sampling which is fundamental in science, engineering, and human affairs. It is also especially important in engineering practice, where things are generally fine if they are going as expected, and you would like to be assured that the probability of an unexpected event is very low.

A final chapter applies the previously probabilitic tools to solve problems involving more complex random processes. You will see why you will probably never get very far ahead at the casino and how two Stanford graduate students became billionaires by combining graph theory and probability theory to design a better search engine for the web.

# 17 Events and Probability Spaces

## 17.1 Let's Make a Deal

In the September 9, 1990 issue of *Parade* magazine, columnist Marilyn vos Savant responded to this letter:

> *Suppose you're on a game show, and you're given the choice of three doors. Behind one door is a car, behind the others, goats. You pick a door, say number 1, and the host, who knows what's behind the doors, opens another door, say number 3, which has a goat. He says to you, "Do you want to pick door number 2?" Is it to your advantage to switch your choice of doors?*
>
> Craig. F. Whitaker
> Columbia, MD

The letter describes a situation like one faced by contestants in the 1970's game show *Let's Make a Deal*, hosted by Monty Hall and Carol Merrill. Marilyn replied that the contestant should indeed switch. She explained that if the car was behind either of the two unpicked doors—which is twice as likely as the the car being behind the picked door—the contestant wins by switching. But she soon received a torrent of letters, many from mathematicians, telling her that she was wrong. The problem became known as the *Monty Hall Problem* and it generated thousands of hours of heated debate.

This incident highlights a fact about probability: the subject uncovers lots of examples where ordinary intuition leads to completely wrong conclusions. So until you've studied probabilities enough to have refined your intuition, a way to avoid errors is to fall back on a rigorous, systematic approach such as the Four Step Method that we will describe shortly. First, let's make sure we really understand the setup for this problem. This is always a good thing to do when you are dealing with probability.

### 17.1.1 Clarifying the Problem

Craig's original letter to Marilyn vos Savant is a bit vague, so we must make some assumptions in order to have any hope of modeling the game formally. For example, we will assume that:

1. The car is equally likely to be hidden behind each of the three doors.
2. The player is equally likely to pick each of the three doors, regardless of the car's location.
3. After the player picks a door, the host *must* open a different door with a goat behind it and offer the player the choice of staying with the original door or switching.
4. If the host has a choice of which door to open, then he is equally likely to select each of them.

In making these assumptions, we're reading a lot into Craig Whitaker's letter. There are other plausible interpretations that lead to different answers. But let's accept these assumptions for now and address the question, "What is the probability that a player who switches wins the car?"

## 17.2 The Four Step Method

Every probability problem involves some sort of randomized experiment, process, or game. And each such problem involves two distinct challenges:

1. How do we model the situation mathematically?
2. How do we solve the resulting mathematical problem?

In this section, we introduce a four step approach to questions of the form, "What is the probability that...?" In this approach, we build a probabilistic model step-by-step, formalizing the original question in terms of that model. Remarkably, the structured thinking that this approach imposes provides simple solutions to many famously-confusing problems. For example, as you'll see, the four step method cuts through the confusion surrounding the Monty Hall problem like a Ginsu knife.

### 17.2.1 Step 1: Find the Sample Space

Our first objective is to identify all the possible outcomes of the experiment. A typical experiment involves several randomly-determined quantities. For example, the Monty Hall game involves three such quantities:

1. The door concealing the car.
2. The door initially chosen by the player.

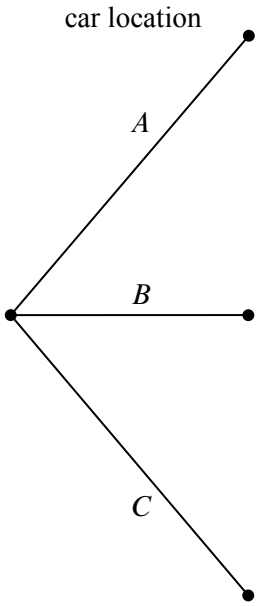



**Figure 17.1** The first level in a tree diagram for the Monty Hall Problem. The branches correspond to the door behind which the car is located.

3. The door that the host opens to reveal a goat.

Every possible combination of these randomly-determined quantities is called an *outcome*. The set of all possible outcomes is called the *sample space* for the experiment.

A *tree diagram* is a graphical tool that can help us work through the four step approach when the number of outcomes is not too large or the problem is nicely structured. In particular, we can use a tree diagram to help understand the sample space of an experiment. The first randomly-determined quantity in our experiment is the door concealing the prize. We represent this as a tree with three branches, as shown in Figure 17.1. In this diagram, the doors are called $A$, $B$, and $C$ instead of 1, 2, and 3, because we'll be adding a lot of other numbers to the picture later.

For each possible location of the prize, the player could initially choose any of the three doors. We represent this in a second layer added to the tree. Then a third layer represents the possibilities of the final step when the host opens a door to reveal a goat, as shown in Figure 17.2.

Notice that the third layer reflects the fact that the host has either one choice or two, depending on the position of the car and the door initially selected by the player. For example, if the prize is behind door A and the player picks door B, then

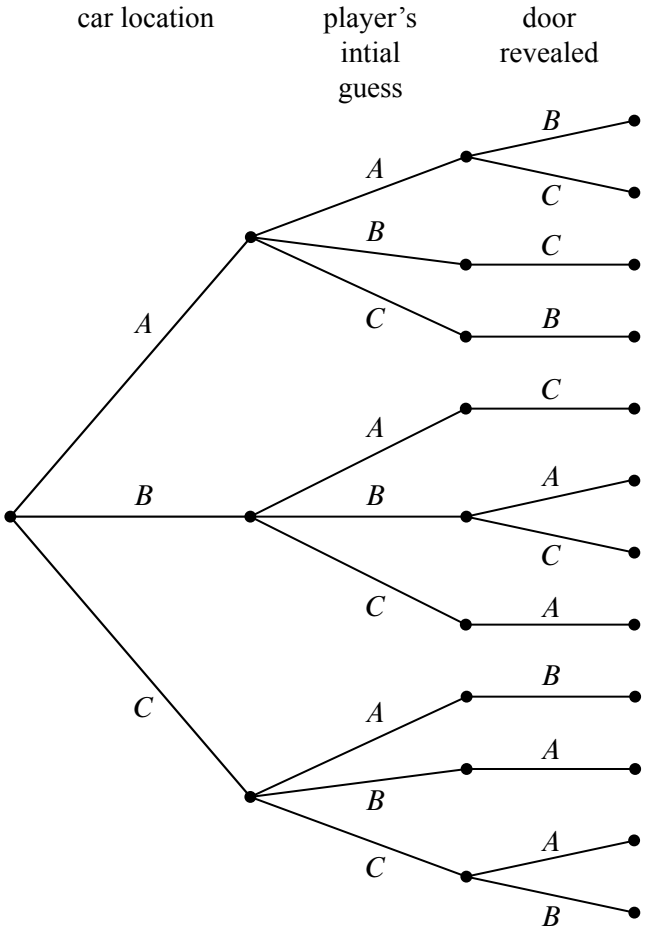



**Figure 17.2** The full tree diagram for the Monty Hall Problem. The second level indicates the door initially chosen by the player. The third level indicates the door revealed by Monty Hall.

the host must open door C. However, if the prize is behind door A and the player picks door A, then the host could open either door B or door C.

Now let's relate this picture to the terms we introduced earlier: the leaves of the tree represent *outcomes* of the experiment, and the set of all leaves represents the *sample space*. Thus, for this experiment, the sample space consists of 12 outcomes. For reference, we've labeled each outcome in Figure 17.3 with a triple of doors indicating:

(door concealing prize, door initially chosen, door opened to reveal a goat).

In these terms, the sample space is the set

$$\mathcal{S} = \left\{ \begin{array}{l} (A,A,B),\ (A,A,C),\ (A,B,C),\ (A,C,B),\ (B,A,C),\ (B,B,A), \\ (B,B,C),\ (B,C,A),\ (C,A,B),\ (C,B,A),\ (C,C,A),\ (C,C,B) \end{array} \right\}$$

The tree diagram has a broader interpretation as well: we can regard the whole experiment as following a path from the root to a leaf, where the branch taken at each stage is "randomly" determined. Keep this interpretation in mind; we'll use it again later.

### 17.2.2 Step 2: Define Events of Interest

Our objective is to answer questions of the form "What is the probability that ...?", where, for example, the missing phrase might be "the player wins by switching", "the player initially picked the door concealing the prize", or "the prize is behind door C." Each of these phrases characterizes a set of outcomes. For example, the outcomes specified by "the prize is behind door $C$" is:

$$\{(C,A,B),(C,B,A),(C,C,A),(C,C,B)\}.$$

A set of outcomes is called an *event* and it is a subset of the sample space. So the event that the player initially picked the door concealing the prize is the set:

$$\{(A,A,B),(A,A,C),(B,B,A),(B,B,C),(C,C,A),(C,C,B)\}.$$

And what we're really after, the event that the player wins by switching, is the set of outcomes:

[switching-wins]

$$::= \{(A,B,C),(A,C,B),(B,A,C),(B,C,A),(C,A,B),(C,B,A)\}. \quad (17.1)$$

These outcomes have check marks in Figure 17.4.

Notice that exactly half of the outcomes are checked, meaning that the player wins by switching in half of all outcomes. You might be tempted to conclude that a player who switches wins with probability $1/2$. *This is wrong.* The reason is that these outcomes are not all equally likely, as we'll see shortly.

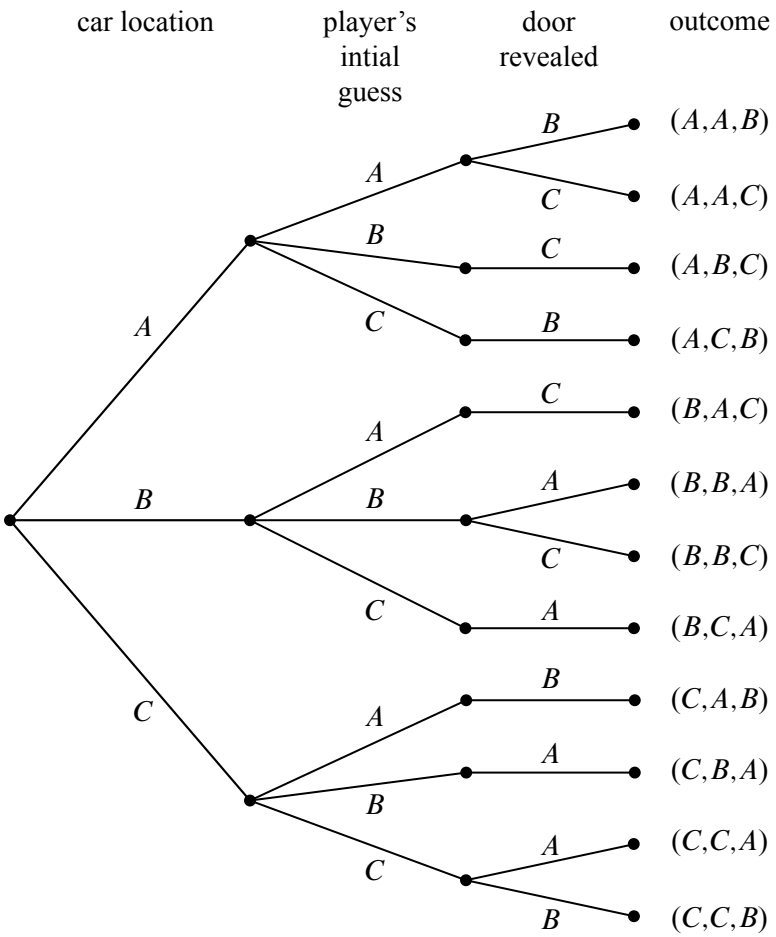



**Figure 17.3** The tree diagram for the Monty Hal Problem with the outcomes labeled for each path from root to leaf. For example, outcome $(A, A, B)$ corresponds to the car being behind door $A$, the player initially choosing door $A$, and Monty Hall revealing the goat behind door $B$.

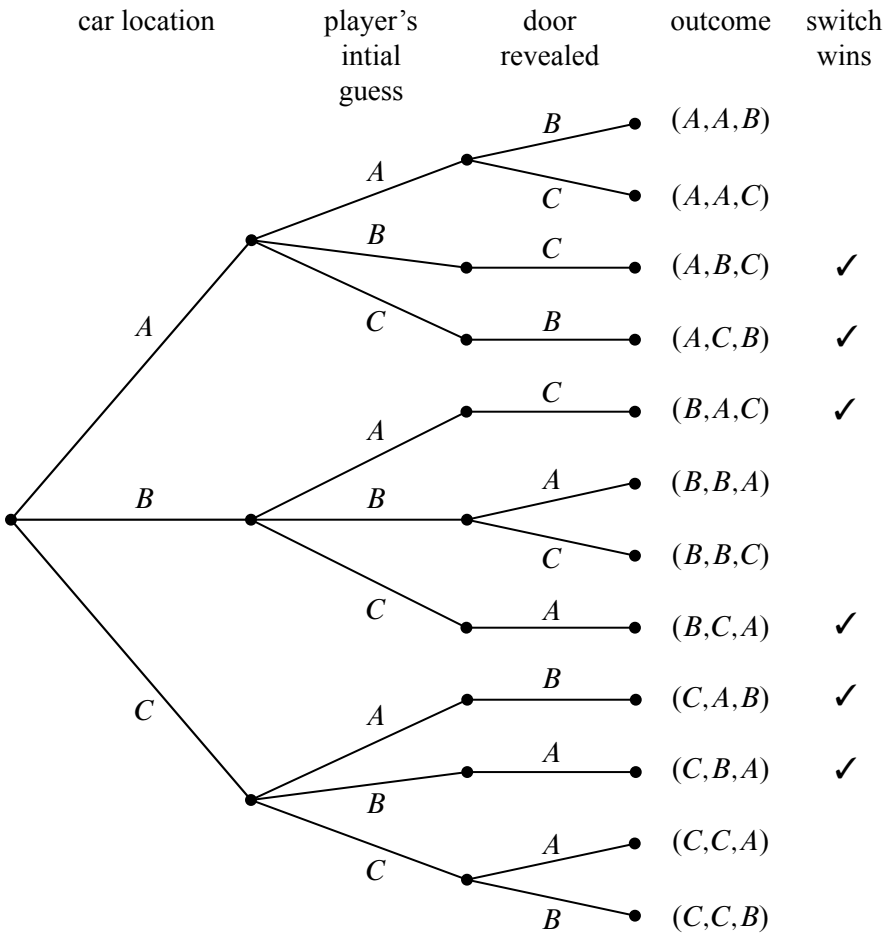



**Figure 17.4** The tree diagram for the Monty Hall Problem where the outcomes in the event where the player wins by switching are denoted with a check mark.

### 17.2.3 Step 3: Determine Outcome Probabilities

So far we've enumerated all the possible outcomes of the experiment. Now we must start assessing the likelihood of those outcomes. In particular, the goal of this step is to assign each outcome a probability, indicating the fraction of the time this outcome is expected to occur. The sum of all outcome probabilities must be one, reflecting the fact that there always is an outcome.

Ultimately, outcome probabilities are determined by the phenomenon we're modeling and thus are not quantities that we can derive mathematically. However, mathematics can help us compute the probability of every outcome *based on fewer and more elementary modeling decisions.* In particular, we'll break the task of determining outcome probabilities into two stages.

#### Step 3a: Assign Edge Probabilities

First, we record a probability on each *edge* of the tree diagram. These edge-probabilities are determined by the assumptions we made at the outset: that the prize is equally likely to be behind each door, that the player is equally likely to pick each door, and that the host is equally likely to reveal each goat, if he has a choice. Notice that when the host has no choice regarding which door to open, the single branch is assigned probability 1. For example, see Figure 17.5.

#### Step 3b: Compute Outcome Probabilities

Our next job is to convert edge probabilities into outcome probabilities. This is a purely mechanical process:

> the probability of an outcome is equal to the product of the edge-probabilities on the path from the root to that outcome.

For example, the probability of the topmost outcome in Figure 17.5, $(A, A, B)$, is

$$\frac{1}{3} \cdot \frac{1}{3} \cdot \frac{1}{2} = \frac{1}{18}.$$

There's an easy, intuitive justification for this rule. As the steps in an experiment progress randomly along a path from the root of the tree to a leaf, the probabilities on the edges indicate how likely the path is to proceed along each branch. For example, a path starting at the root in our example is equally likely to go down each of the three top-level branches.

How likely is such a path to arrive at the topmost outcome, $(A, A, B)$? Well, there is a 1-in-3 chance that a path would follow the $A$-branch at the top level, a 1-in-3 chance it would continue along the $A$-branch at the second level, and 1-in-2 chance it would follow the $B$-branch at the third level. Thus, it seems that

1 path in 18 should arrive at the $(A, A, B)$ leaf, which is precisely the probability we assign it.

We have illustrated all of the outcome probabilities in Figure 17.5.

Specifying the probability of each outcome amounts to defining a function that maps each outcome to a probability. This function is usually called $\Pr[\cdot]$. In these terms, we've just determined that:

$$\begin{aligned}\Pr[(A, A, B)] &= \frac{1}{18},\\ \Pr[(A, A, C)] &= \frac{1}{18},\\ \Pr[(A, B, C)] &= \frac{1}{9},\\ &\text{etc.}\end{aligned}$$

### 17.2.4 Step 4: Compute Event Probabilities

We now have a probability for each *outcome*, but we want to determine the probability of an *event*. The probability of an event $E$ is denoted by $\Pr[E]$ and it is the sum of the probabilities of the outcomes in $E$. For example, the probability of the [switching wins] event (17.1) is

$$\begin{aligned}&\Pr[\text{[switching wins]}]\\ &= \Pr[(A, B, C)] + \Pr[(A, C, B)] + \Pr[(B, A, C)] +\\ &\qquad \Pr[(B, C, A)] + \Pr[(C, A, B)] + \Pr[(C, B, A)]\\ &= \frac{1}{9} + \frac{1}{9} + \frac{1}{9} + \frac{1}{9} + \frac{1}{9} + \frac{1}{9}\\ &= \frac{2}{3}.\end{aligned}$$

It seems Marilyn's answer is correct! A player who switches doors wins the car with probability $2/3$. In contrast, a player who stays with his or her original door wins with probability $1/3$, since staying wins if and only if switching loses.

We're done with the problem! We didn't need any appeals to intuition or ingenious analogies. In fact, no mathematics more difficult than adding and multiplying fractions was required. The only hard part was resisting the temptation to leap to an "intuitively obvious" answer.

### 17.2.5 An Alternative Interpretation of the Monty Hall Problem

Was Marilyn really right? Our analysis indicates that she was. But a more accurate conclusion is that her answer is correct *provided we accept her interpretation of the*

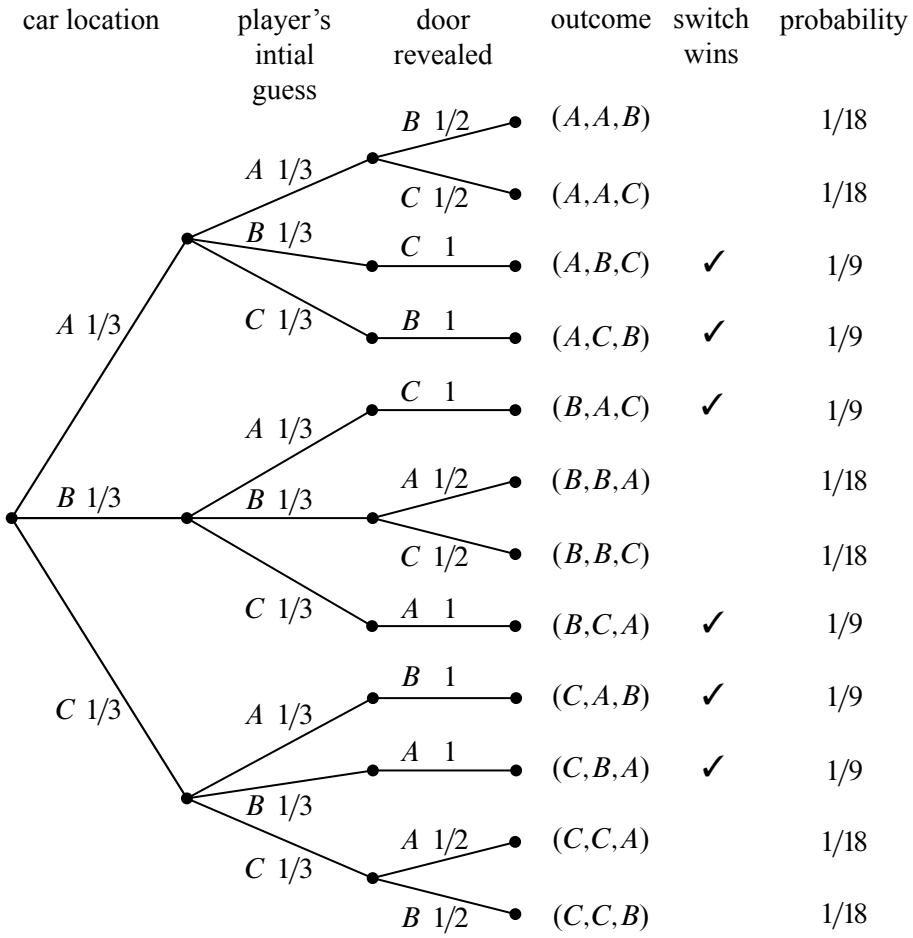
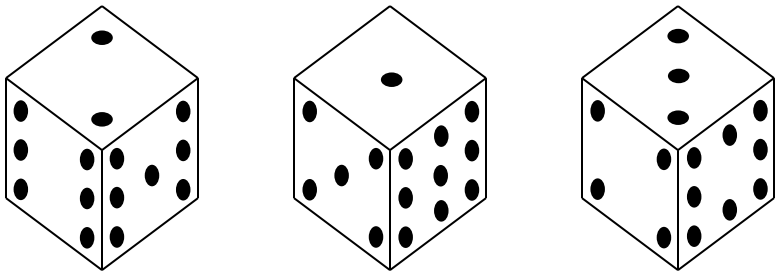



**Figure 17.5** The tree diagram for the Monty Hall Problem where edge weights denote the probability of that branch being taken given that we are at the parent of that branch. For example, if the car is behind door $A$, then there is a 1/3 chance that the player's initial selection is door $B$. The rightmost column shows the outcome probabilities for the Monty Hall Problem. Each outcome probability is simply the product of the probabilities on the path from the root to the outcome leaf.



*A* *B* *C*

**Figure 17.6** The strange dice. The number of pips on each concealed face is the same as the number on the opposite face. For example, when you roll die $A$, the probabilities of getting a 2, 6, or 7 are each $1/3$.

*question*. There is an equally plausible interpretation in which Marilyn's answer is wrong. Notice that Craig Whitaker's original letter does not say that the host is *required* to reveal a goat and offer the player the option to switch, merely that he *did* these things. In fact, on the *Let's Make a Deal* show, Monty Hall sometimes simply opened the door that the contestant picked initially. Therefore, if he wanted to, Monty could give the option of switching only to contestants who picked the correct door initially. In this case, switching never works!

## 17.3 Strange Dice

The four-step method is surprisingly powerful. Let's get some more practice with it. Imagine, if you will, the following scenario.

It's a typical Saturday night. You're at your favorite pub, contemplating the true meaning of infinite cardinalities, when a burly-looking biker plops down on the stool next to you. Just as you are about to get your mind around $\mathcal{P}(\mathcal{P}(\mathbb{R}))$, biker dude slaps three strange-looking dice on the bar and challenges you to a \$100 wager. His rules are simple. Each player selects one die and rolls it once. The player with the lower value pays the other player \$100.

Naturally, you are skeptical, especially after you see that these are not ordinary dice. Each die has the usual six sides, but opposite sides have the same number on them, and the numbers on the dice are different, as shown in Figure 17.6.

Biker dude notices your hesitation, so he sweetens his offer: he will pay you \$105 if you roll the higher number, but you only need pay him \$100 if he rolls

higher, *and* he will let you pick a die first, after which he will pick one of the other two. The sweetened deal sounds persuasive since it gives you a chance to pick what you think is the best die, so you decide you will play. But which of the dice should you choose? Die $B$ is appealing because it has a 9, which is a sure winner if it comes up. Then again, die $A$ has two fairly large numbers and die $C$ has an 8 and no really small values.

In the end, you choose die $B$ because it has a 9, and then biker dude selects die $A$. Let's see what the probability is that you will win. (Of course, you probably should have done this before picking die $B$ in the first place.) Not surprisingly, we will use the four-step method to compute this probability.

### 17.3.1 Die $A$ versus Die $B$

***Step 1: Find the sample space.***

The tree diagram for this scenario is shown in Figure 17.7. In particular, the sample space for this experiment are the nine pairs of values that might be rolled with Die $A$ and Die $B$:

For this experiment, the sample space is a set of nine outcomes:

$$\mathcal{S} = \{(2,1),\ (2,5),\ (2,9),\ (6,1),\ (6,5),\ (6,9),\ (7,1),\ (7,5),\ (7,9)\}.$$

***Step 2: Define events of interest.***

We are interested in the event that the number on die $A$ is greater than the number on die $B$. This event is a set of five outcomes:

$$\{(2,1),\ (6,1),\ (6,5),\ (7,1),\ (7,5)\}.$$

These outcomes are marked $A$ in the tree diagram in Figure 17.7.

***Step 3: Determine outcome probabilities.***

To find outcome probabilities, we first assign probabilities to edges in the tree diagram. Each number on each die comes up with probability $1/3$, regardless of the value of the other die. Therefore, we assign all edges probability $1/3$. The probability of an outcome is the product of the probabilities on the corresponding root-to-leaf path, which means that every outcome has probability $1/9$. These probabilities are recorded on the right side of the tree diagram in Figure 17.7.

***Step 4: Compute event probabilities.***

The probability of an event is the sum of the probabilities of the outcomes in that event. In this case, all the outcome probabilities are the same, so we say that the sample space is *uniform*. Computing event probabilities for uniform sample spaces

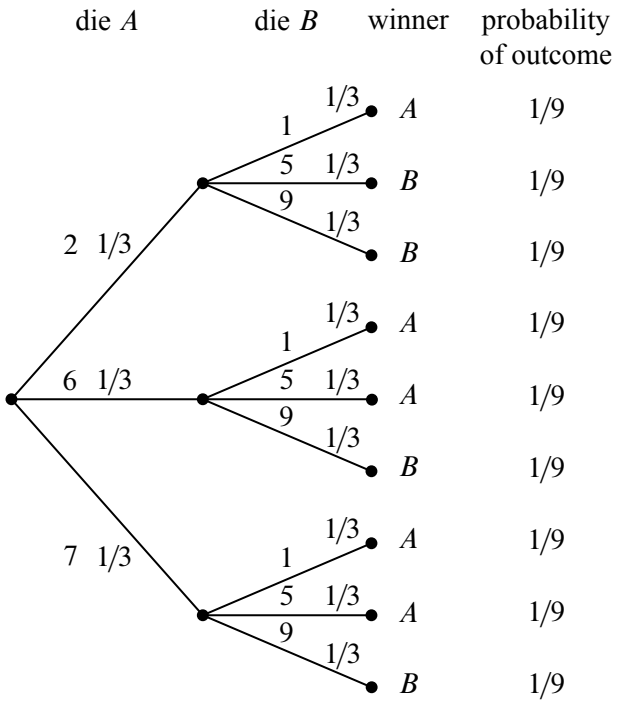



**Figure 17.7** The tree diagram for one roll of die $A$ versus die $B$. Die $A$ wins with probability $5/9$.

is particularly easy since you just have to compute the number of outcomes in the event. In particular, for any event $E$ in a uniform sample space $\mathcal{S}$,

$$\Pr[E] = \frac{|E|}{|\mathcal{S}|}. \tag{17.2}$$

In this case, $E$ is the event that die $A$ beats die $B$, so $|E| = 5$, $|\mathcal{S}| = 9$, and

$$\Pr[E] = 5/9.$$

This is bad news for you. Die $A$ beats die $B$ more than half the time and, not surprisingly, you just lost \$100.

Biker dude consoles you on your "bad luck" and, given that he's a sensitive guy beneath all that leather, he offers to go double or nothing.[1] Given that your wallet only has \$25 in it, this sounds like a good plan. Plus, you figure that choosing die $A$ will give *you* the advantage.

So you choose $A$, and then biker dude chooses $C$. Can you guess who is more likely to win? (Hint: it is generally not a good idea to gamble with someone you don't know in a bar, especially when you are gambling with strange dice.)

### 17.3.2 Die $A$ versus Die $C$

We can construct the three diagram and outcome probabilities as before. The result is shown in Figure 17.8 and there is bad news again. Die $C$ will beat die $A$ with probability $5/9$, and you lose once again.

You now owe the biker dude \$200 and he asks for his money. You reply that you need to go to the bathroom.

### 17.3.3 Die $B$ versus Die $C$

Being a sensitive guy, biker dude nods understandingly and offers yet another wager. This time, he'll let you have die $C$. He'll even let you raise the wager to \$200 so you can win your money back.

This is too good a deal to pass up. You know that die $C$ is likely to beat die $A$ and that die $A$ is likely to beat die $B$, and so die $C$ is *surely* the best. Whether biker dude picks $A$ or $B$, the odds would be in your favor this time. Biker dude must really be a nice guy.

So you pick $C$, and then biker dude picks $B$. Wait, how come you haven't caught on yet and worked out the tree diagram before you took this bet :-) ? If

[1] *Double or nothing* is slang for doing another wager after you have lost the first. If you lose again, you will owe biker dude *double* what you owed him before. If you win, you will owe him *nothing*; in fact, since he should pay you \$210 if he loses, you would come out \$10 ahead.

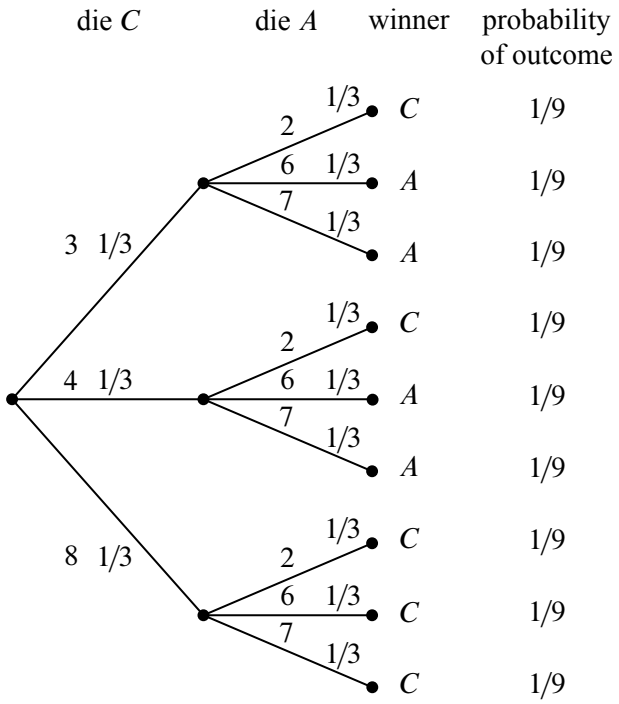



**Figure 17.8** The tree diagram for one roll of die $C$ versus die $A$. Die $C$ wins with probability $5/9$.

you do it now, you'll see by the same reasoning as before that $B$ beats $C$ with probability $5/9$. But surely there is a mistake! How is it possible that

$C$ beats $A$ with probability $5/9$,

$A$ beats $B$ with probability $5/9$,

$B$ beats $C$ with probability $5/9$?

The problem is not with the math, but with your intuition. Since $A$ will beat $B$ more often than not, and $B$ will beat $C$ more often than not, it *seems* like $A$ ought to beat $C$ more often than not, that is, the "beats more often" relation ought to be *transitive*. But this intuitive idea is simply false: whatever die you pick, biker dude can pick one of the others and be likely to win. So picking first is actually a big disadvantage, and as a result, you now owe biker dude $400.

Just when you think matters can't get worse, biker dude offers you one final wager for $1,000. This time, instead of rolling each die once, you will each roll your die twice, and your score is the sum of your rolls, and he will even let you pick your die second, that is, after he picks his. Biker dude chooses die $B$. Now you know that die $A$ will beat die $B$ with probability $5/9$ on one roll, so, jumping at this chance to get ahead, you agree to play, and you pick die $A$. After all, you figure that since a roll of die $A$ beats a roll of die $B$ more often that not, two rolls of die $A$ are even more likely to beat two rolls of die $B$, right?

Wrong! (Did we mention that playing strange gambling games with strangers in a bar is a bad idea?)

### 17.3.4 Rolling Twice

If each player rolls twice, the tree diagram will have four levels and $3^4 = 81$ outcomes. This means that it will take a while to write down the entire tree diagram. But it's easy to write down the first two levels as in Figure 17.9(a) and then notice that the remaining two levels consist of nine identical copies of the tree in Figure 17.9(b).

The probability of each outcome is $(1/3)^4 = 1/81$ and so, once again, we have a uniform probability space. By equation (17.2), this means that the probability that $A$ wins is the number of outcomes where $A$ beats $B$ divided by 81.

To compute the number of outcomes where $A$ beats $B$, we observe that the sum of the two rolls of die $A$ is equally likely to be any element of the following multiset:

$$S_A = \{4, 8, 8, 9, 9, 12, 13, 13, 14\}.$$

The sum of two rolls of die $B$ is equally likely to be any element of the following multiset:

$$S_B = \{2, 6, 6, 10, 10, 10, 14, 14, 18\}.$$

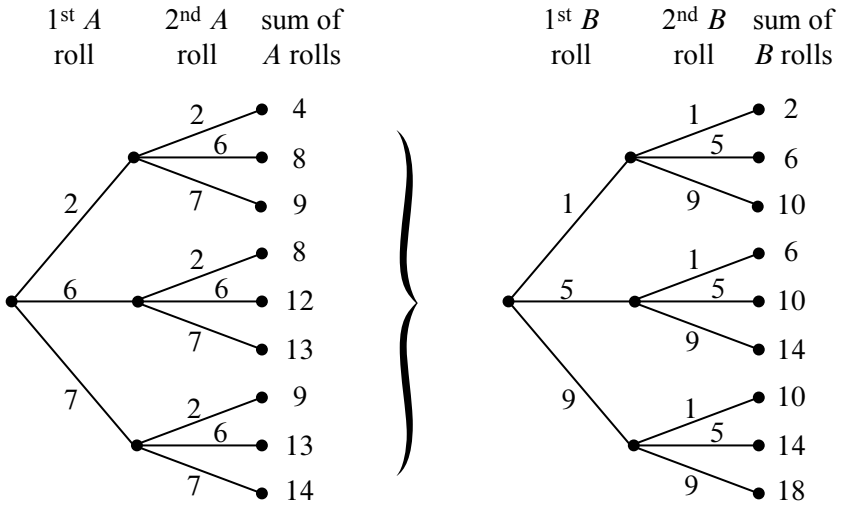



**Figure 17.9** Parts of the tree diagram for die $B$ versus die $A$ where each die is rolled twice. The first two levels are shown in (a). The last two levels consist of nine copies of the tree in (b).

We can treat each outcome as a pair $(x, y) \in \mathcal{S}_A \times \mathcal{S}_B$, where $A$ wins iff $x > y$. If $x = 4$, there is only one $y$ (namely $y = 2$) for which $x > y$. If $x = 8$, there are three values of $y$ for which $x > y$. Continuing the count in this way, the number of pairs for which $x > y$ is

$$1 + 3 + 3 + 3 + 3 + 6 + 6 + 6 + 6 = 37.$$

A similar count shows that there are 42 pairs for which $x < y$, and there are two pairs ((14, 14), (14, 14)) which result in ties. This means that $A$ *loses* to $B$ with probability $42/81 > 1/2$ and ties with probability $2/81$. Die $A$ wins with probability only $37/81$.

How can it be that $A$ is more likely than $B$ to win with one roll, but $B$ is more likely to win with two rolls? Well, why not? The only reason we'd think otherwise is our unreliable, untrained intuition. (Even the authors were surprised when they first learned about this, but at least we didn't lose $1400 to biker dude. :-) ) In fact, the die strength reverses no matter which two die we picked. So for 1 roll,

$$A \succ B \succ C \succ A,$$

but for two rolls,

$$A \prec B \prec C \prec A,$$

where we have used the symbols $\succ$ and $\prec$ to denote which die is more likely to result in the larger value.

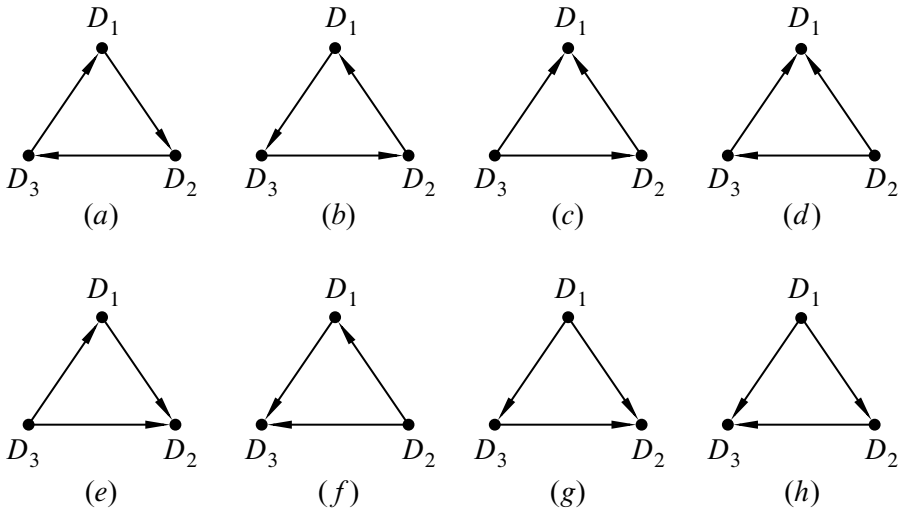



**Figure 17.10** All possible relative strengths for three dice $D_1$, $D_2$, and $D_3$. The edge $\langle D_i \to D_j \rangle$ denotes that the sum of rolls for $D_i$ is likely to be greater than the sum of rolls for $D_j$.

### Even Stranger Dice

The weird behavior of the three strange dice above generalizes in a remarkable way.[2] The idea is that you can find arbitrarily large sets of dice which will beat each other in any desired pattern according to how many times the dice are rolled. The precise statement of this result involves several alternations of universal and existential quantifiers, so it may take a few readings to understand what it is saying:

**Theorem 17.3.1.** *For any $n \geq 2$, there is a set of $n$ dice with the following property: for* any *$n$-node digraph with exactly one directed edge between every two distinct nodes,*[3] *there is a number of rolls $k$ such that the sum of $k$ rolls of the $i$th die is bigger than the sum for the $j$th die with probability greater than $1/2$ iff there is an edge from the $i$th to the $j$th node in the graph.*

For example, the eight possible relative strengths for $n = 3$ dice are shown in Figure 17.10.

Our analysis for the dice in Figure 17.6 showed that for 1 roll, we have the relative strengths shown in Figure 17.10(a), and for two rolls, we have the (reverse) relative strengths shown in Figure 17.10(b). If you are prone to gambling with

[2] Reference Ron Graham paper.

[3] In other words, for every pair of nodes $u \neq v$, either $\langle u \to v \rangle$ or $\langle v \to u \rangle$, but not both, are edges of the graph. Such graphs are called *tournament graphs*, see Problem 9.5.

strangers in bars, it would be a good idea to try figuring out what other relative strengths are possible for the dice in Figure 17.6 when using more rolls.

## 17.4 Set Theory and Probability

Let's abstract what we've just done with the Monty Hall and strange dice examples into a general mathematical definition of sample spaces and probability.

### 17.4.1 Probability Spaces

**Definition 17.4.1.** A countable *sample space* $\mathcal{S}$ is a nonempty countable set.[4] An element $\omega \in \mathcal{S}$ is called an *outcome*. A subset of $\mathcal{S}$ is called an *event*.

**Definition 17.4.2.** A *probability function* on a sample space $\mathcal{S}$ is a total function $\Pr : \mathcal{S} \to \mathbb{R}$ such that

- $\Pr[\omega] \geq 0$ for all $\omega \in \mathcal{S}$, and
- $\sum_{\omega \in \mathcal{S}} \Pr[\omega] = 1$.

A sample space together with a probability function is called a *probability space*. For any event $E \subseteq \mathcal{S}$, the *probability of* $E$ is defined to be the sum of the probabilities of the outcomes in $E$:

$$\Pr[E] ::= \sum_{\omega \in E} \Pr[\omega].$$

In the previous examples there were only finitely many possible outcomes, but we'll quickly come to examples that have a countably infinite number of outcomes.

The study of probability is closely tied to set theory because any set can be a sample space and any subset can be an event. General probability theory deals with uncountable sets like the set of real numbers, but we won't need these, and sticking to countable sets lets us define the probability of events using sums instead of integrals. It also lets us avoid some distracting technical problems in set theory like the Banach-Tarski "paradox" mentioned in Chapter 5.

### 17.4.2 Probability Rules from Set Theory

Most of the rules and identities that we have developed for finite sets extend very naturally to probability.

[4] Yes, sample spaces can be infinite. If you did not read Chapter 5, don't worry —*countable* just means that you can list the elements of the sample space as $\omega_0, \omega_1, \omega_2, \ldots$.

An immediate consequence of the definition of event probability is that for *disjoint* events $E$ and $F$,

$$\Pr[E \cup F] = \Pr[E] + \Pr[F].$$

This generalizes to a countable number of events, as follows.

**Rule 17.4.3** (Sum Rule)**.** *If $\{E_0, E_1, \ldots\}$ is collection of disjoint events, then*

$$\Pr\left[\bigcup_{n\in\mathbb{N}} E_n\right] = \sum_{n\in\mathbb{N}} \Pr[E_n].$$

The Sum Rule lets us analyze a complicated event by breaking it down into simpler cases. For example, if the probability that a randomly chosen MIT student is native to the United States is 60%, to Canada is 5%, and to Mexico is 5%, then the probability that a random MIT student is native to North America is 70%.

Another consequence of the Sum Rule is that $\Pr[A] + \Pr[\overline{A}] = 1$, which follows because $\Pr[\mathcal{S}] = 1$ and $\mathcal{S}$ is the union of the disjoint sets $A$ and $\overline{A}$. This equation often comes up in the form:

$$\Pr[\overline{A}] = 1 - \Pr[A]. \qquad \text{(Complement Rule)}$$

Sometimes the easiest way to compute the probability of an event is to compute the probability of its complement and then apply this formula.

Some further basic facts about probability parallel facts about cardinalities of finite sets. In particular:

$$\begin{aligned}
&\Pr[B - A] = \Pr[B] - \Pr[A \cap B], && \text{(Difference Rule)}\\
&\Pr[A \cup B] = \Pr[A] + \Pr[B] - \Pr[A \cap B], && \text{(Inclusion-Exclusion)}\\
&\Pr[A \cup B] \le \Pr[A] + \Pr[B], && \text{(Boole's Inequality)}\\
&\text{If } A \subseteq B, \text{ then } \Pr[A] \le \Pr[B]. && \text{(Monotonicity Rule)}
\end{aligned}$$

The Difference Rule follows from the Sum Rule because $B$ is the union of the disjoint sets $B - A$ and $A \cap B$. Inclusion-Exclusion then follows from the Sum and Difference Rules, because $A \cup B$ is the union of the disjoint sets $A$ and $B - A$. Boole's inequality is an immediate consequence of Inclusion-Exclusion since probabilities are nonnegative. Monotonicity follows from the definition of event probability and the fact that outcome probabilities are nonnegative.

The two-event Inclusion-Exclusion equation above generalizes to $n$ events in the same way as the corresponding Inclusion-Exclusion rule for $n$ sets. Boole's inequality also generalizes to

**Rule 17.4.4** (Union Bound)**.**

$$\Pr[E_1 \cup \cdots \cup E_n] \leq \Pr[E_1] + \cdots + \Pr[E_n]. \tag{17.3}$$

This simple Union Bound is useful in many calculations. For example, suppose that $E_i$ is the event that the $i$-th critical component in a spacecraft fails. Then $E_1 \cup \cdots \cup E_n$ is the event that *some* critical component fails. If $\sum_{i=1}^{n} \Pr[E_i]$ is small, then the Union Bound can give an adequate upper bound on this vital probability.

### 17.4.3 Uniform Probability Spaces

**Definition 17.4.5.** A finite probability space, $\mathcal{S}$, is said to be *uniform* if $\Pr[\omega]$ is the same for every outcome $\omega \in \mathcal{S}$.

As we saw in the strange dice problem, uniform sample spaces are particularly easy to work with. That's because for any event $E \subseteq \mathcal{S}$,

$$\Pr[E] = \frac{|E|}{|\mathcal{S}|}. \tag{17.4}$$

This means that once we know the cardinality of $E$ and $\mathcal{S}$, we can immediately obtain $\Pr[E]$. That's great news because we developed lots of tools for computing the cardinality of a set in Part III.

For example, suppose that you select five cards at random from a standard deck of 52 cards. What is the probability of having a full house? Normally, this question would take some effort to answer. But from the analysis in Section 15.9.2, we know that

$$|\mathcal{S}| = \binom{52}{5}$$

and

$$|E| = 13 \cdot \binom{4}{3} \cdot 12 \cdot \binom{4}{2}$$

where $E$ is the event that we have a full house. Since every five-card hand is equally likely, we can apply equation (17.4) to find that

$$\begin{aligned} \Pr[E] &= \frac{13 \cdot 12 \cdot \binom{4}{3} \cdot \binom{4}{2}}{\binom{52}{5}} \\ &= \frac{13 \cdot 12 \cdot 4 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2}{52 \cdot 51 \cdot 50 \cdot 49 \cdot 48} = \frac{18}{12495} \\ &\approx \frac{1}{694}. \end{aligned}$$

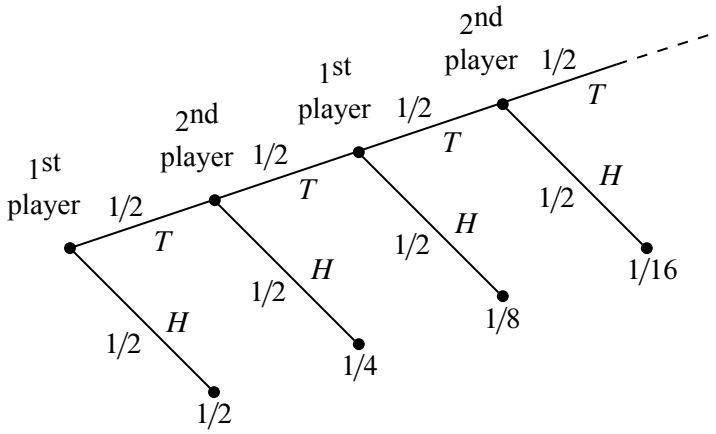



**Figure 17.11** The tree diagram for the game where players take turns flipping a fair coin. The first player to flip heads wins.

### 17.4.4 Infinite Probability Spaces

Infinite probability spaces are fairly common. For example, two players take turns flipping a fair coin. Whoever flips heads first is declared the winner. What is the probability that the first player wins? A tree diagram for this problem is shown in Figure 17.11.

The event that the first player wins contains an infinite number of outcomes, but we can still sum their probabilities:

$$\begin{aligned}\Pr[\text{first player wins}] &= \frac{1}{2} + \frac{1}{8} + \frac{1}{32} + \frac{1}{128} + \cdots \\ &= \frac{1}{2}\sum_{n=0}^{\infty}\left(\frac{1}{4}\right)^n \\ &= \frac{1}{2}\left(\frac{1}{1-1/4}\right) = \frac{2}{3}.\end{aligned}$$

Similarly, we can compute the probability that the second player wins:

$$\Pr[\text{second player wins}] = \frac{1}{4} + \frac{1}{16} + \frac{1}{64} + \frac{1}{256} + \cdots = \frac{1}{3}.$$

In this case, the sample space is the infinite set

$$\mathcal{S} ::= \{\, \mathrm{T}^n\mathrm{H} \mid n \in \mathbb{N} \,\},$$

where $T^n$ stands for a length $n$ string of T's. The probability function is

$$\Pr[T^n H] ::= \frac{1}{2^{n+1}}.$$

To verify that this is a probability space, we just have to check that all the probabilities are nonnegative and that they sum to 1. Nonnegativity is obvious, and applying the formula for the sum of a geometric series, we find that

$$\sum_{n\in\mathbb{N}} \Pr[T^n H] = \sum_{n\in\mathbb{N}} \frac{1}{2^{n+1}} = 1.$$

Notice that this model does not have an outcome corresponding to the possibility that both players keep flipping tails forever —in the diagram, flipping forever corresponds to following the infinite path in the tree without ever reaching a leaf/outcome. If leaving this possibility out of the model bothers you, you're welcome to fix it by adding another outcome, $\omega_{\text{forever}}$, to indicate that that's what happened. Of course since the probabililities of the other outcomes already sum to 1, you have to define the probability of $\omega_{\text{forever}}$ to be 0. Now outcomes with probability zero will have no impact on our calculations, so there's no harm in adding it in if it makes you happier. On the other hand, in countable probability spaces it isn't necessary to have outcomes with probability zero, and we will generally ignore them.

## 17.5 Conditional Probability

Suppose that we pick a random person in the world. Everyone has an equal chance of being selected. Let $A$ be the event that the person is an MIT student, and let $B$ be the event that the person lives in Cambridge. What are the probabilities of these events? Intuitively, we're picking a random point in the big ellipse shown in Figure 17.12 and asking how likely that point is to fall into region $A$ or $B$.

The vast majority of people in the world neither live in Cambridge nor are MIT students, so events $A$ and $B$ both have low probability. But what about the probability that a person is an MIT student, *given* that the person lives in Cambridge? This should be much greater —but what is it exactly?

What we're asking for is called a *conditional probability*; that is, the probability that one event happens, given that some other event definitely happens. Questions about conditional probabilities come up all the time:

- What is the probability that it will rain this afternoon, given that it is cloudy this morning?

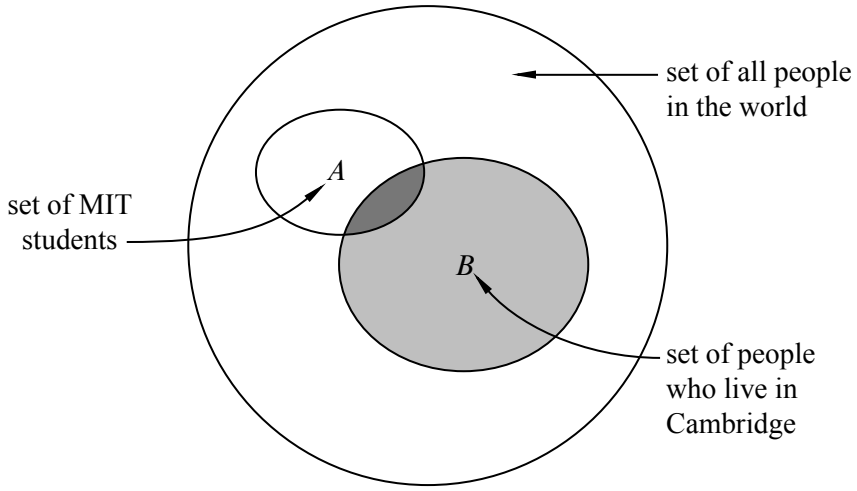



**Figure 17.12** Selecting a random person. $A$ is the event that the person is an MIT student. $B$ is the event that the person lives in Cambridge.

- What is the probability that two rolled dice sum to 10, given that both are odd?
- What is the probability that I'll get four-of-a-kind in Texas No Limit Hold 'Em Poker, given that I'm initially dealt two queens?

There is a special notation for conditional probabilities. In general, $\Pr[A \mid B]$ denotes the probability of event $A$, given that event $B$ happens. So, in our example, $\Pr[A \mid B]$ is the probability that a random person is an MIT student, given that he or she is a Cambridge resident.

How do we compute $\Pr[A \mid B]$? Since we are *given* that the person lives in Cambridge, we can forget about everyone in the world who does not. Thus, all outcomes outside event $B$ are irrelevant. So, intuitively, $\Pr[A \mid B]$ should be the fraction of Cambridge residents that are also MIT students; that is, the answer should be the probability that the person is in set $A \cap B$ (the darkly shaded region in Figure 17.12) divided by the probability that the person is in set $B$ (the lightly shaded region). This motivates the definition of conditional probability:

**Definition 17.5.1.**

$$\Pr[A \mid B] ::= \frac{\Pr[A \cap B]}{\Pr[B]}$$

If $\Pr[B] = 0$, then the conditional probability $\Pr[A \mid B]$ is undefined.

Pure probability is often counterintuitive, but conditional probability is even worse! Conditioning can subtly alter probabilities and produce unexpected results in randomized algorithms and computer systems as well as in betting games. Yet, the mathematical definition of conditional probability given above is very simple and should give you no trouble —provided that you rely on mathematical reasoning and not intuition. The four-step method will also be very helpful as we will see in the next examples.

### 17.5.1 The Four-Step Method for Conditional Probability: The "Halting Problem"

The *Halting Problem* was the first example of a property that could not be tested by any program. It was introduced by Alan Turing in his seminal 1936 paper. The problem is to determine whether a Turing machine halts on a given ... yadda yadda yadda ... more importantly, it was *the name of the MIT EECS department's famed C-league hockey team*.

In a best-of-three tournament, the Halting Problem wins the first game with probability $1/2$. In subsequent games, their probability of winning is determined by the outcome of the previous game. If the Halting Problem won the previous game, then they are invigorated by victory and win the current game with probability $2/3$. If they lost the previous game, then they are demoralized by defeat and win the current game with probability only $1/3$. What is the probability that the Halting Problem wins the tournament, given that they win the first game?

This is a question about a conditional probability. Let $A$ be the event that the Halting Problem wins the tournament, and let $B$ be the event that they win the first game. Our goal is then to determine the conditional probability $\Pr[A \mid B]$.

We can tackle conditional probability questions just like ordinary probability problems: using a tree diagram and the four step method. A complete tree diagram is shown in Figure 17.13.

***Step 1: Find the Sample Space***

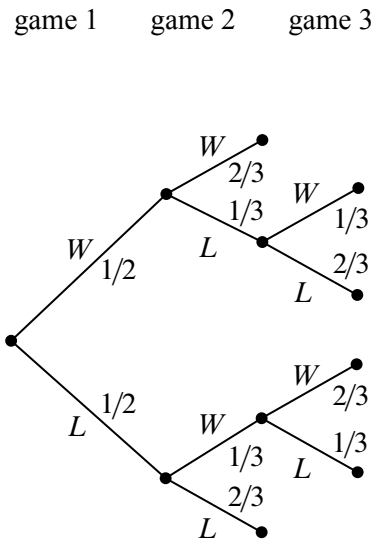
Each internal vertex in the tree diagram has two children, one corresponding to a win for the Halting Problem (labeled $W$) and one corresponding to a loss (labeled $L$). The complete sample space is:

$$S = \{WW,\ WLW,\ WLL,\ LWW,\ LWL,\ LL\}.$$

***Step 2: Define Events of Interest***

The event that the Halting Problem wins the whole tournament is:

$$T = \{WW,\ WLW,\ LWW\}.$$




| game 1 | game 2 | game 3 | outcome | event A: win the series | event B: win game 1 | outcome probability |
|---|---|---|---|---|---|---|
| | | | $WW$ | ✓ | ✓ | 1/3 |
| | | | $WLW$ | ✓ | ✓ | 1/18 |
| | | | $WLL$ | | ✓ | 1/9 |
| | | | $LWW$ | ✓ | | 1/9 |
| | | | $LWL$ | | | 1/18 |
| | | | $LL$ | | | 1/3 |

**Figure 17.13** The tree diagram for computing the probability that the "Halting Problem" wins two out of three games given that they won the first game.

And the event that the Halting Problem wins the first game is:

$$F = \{WW,\ WLW,\ WLL\}.$$

The outcomes in these events are indicated with check marks in the tree diagram in Figure 17.13.

***Step 3: Determine Outcome Probabilities***

Next, we must assign a probability to each outcome. We begin by labeling edges as specified in the problem statement. Specifically, The Halting Problem has a $1/2$ chance of winning the first game, so the two edges leaving the root are each assigned probability $1/2$. Other edges are labeled $1/3$ or $2/3$ based on the outcome of the preceding game. We then find the probability of each outcome by multiplying all probabilities along the corresponding root-to-leaf path. For example, the probability of outcome $WLL$ is:

$$\frac{1}{2} \cdot \frac{1}{3} \cdot \frac{2}{3} = \frac{1}{9}.$$

**Step 4: Compute Event Probabilities**

We can now compute the probability that The Halting Problem wins the tournament, given that they win the first game:

$$\begin{aligned}\Pr\left[A \mid B\right] &= \frac{\Pr[A \cap B]}{\Pr[B]} \\ &= \frac{\Pr[\{WW, WLW\}]}{\Pr[\{WW, WLW, WLL\}]} \\ &= \frac{1/3 + 1/18}{1/3 + 1/18 + 1/9} \\ &= \frac{7}{9}.\end{aligned}$$

We're done! If the Halting Problem wins the first game, then they win the whole tournament with probability $7/9$.

### 17.5.2 Why Tree Diagrams Work

We've now settled into a routine of solving probability problems using tree diagrams. But we've left a big question unaddressed: what is the mathematical justification behind those funny little pictures? Why do they work?

The answer involves conditional probabilities. In fact, the probabilities that we've been recording on the edges of tree diagrams *are* conditional probabilities. For example, consider the uppermost path in the tree diagram for the Halting Problem, which corresponds to the outcome $WW$. The first edge is labeled $1/2$, which is the probability that the Halting Problem wins the first game. The second edge is labeled $2/3$, which is the probability that the Halting Problem wins the second game, *given* that they won the first —that's a conditional probability! More generally, on each edge of a tree diagram, we record the probability that the experiment proceeds along that path, given that it reaches the parent vertex.

So we've been using conditional probabilities all along. But why can we multiply edge probabilities to get outcome probabilities? For example, we concluded that:

$$\Pr[WW] = \frac{1}{2} \cdot \frac{2}{3} = \frac{1}{3}.$$

Why is this correct?

The answer goes back to Definition 17.5.1 of conditional probability which could be written in a form called the *Product Rule* for probabilities:

**Rule** (Product Rule: 2 Events)**.** *If* $\Pr[E_1] \neq 0$*, then:*

$$\Pr[E_1 \cap E_2] = \Pr[E_1] \cdot \Pr\left[E_2 \mid E_1\right].$$

Multiplying edge probabilities in a tree diagram amounts to evaluating the right side of this equation. For example:

$$\begin{aligned}&\Pr[\text{win first game} \cap \text{win second game}]\\ &\quad = \Pr[\text{win first game}] \cdot \Pr\left[\text{win second game} \mid \text{win first game}\right]\\ &\quad = \frac{1}{2} \cdot \frac{2}{3}.\end{aligned}$$

So the Product Rule is the formal justification for multiplying edge probabilities to get outcome probabilities! Of course to justify multiplying edge probabilities along longer paths, we need a Product Rule for $n$ events.

**Rule** (Product Rule: $n$ Events)**.**

$$\begin{aligned}\Pr[E_1 \cap E_2 \cap \ldots \cap E_n] = \Pr[E_1] \cdot \Pr\left[E_2 \mid E_1\right] \cdot \Pr\left[E_3 \mid E_1 \cap E_2\right] \cdots\\ \cdot \Pr\left[E_n \mid E_1 \cap E_2 \cap \ldots \cap E_{n-1}\right]\end{aligned}$$

*provided that*

$$\Pr[E_1 \cap E_2 \cap \cdots \cap E_{n-1}] \neq 0.$$

This rule follows by routine induction from the definition of conditional probability.

### 17.5.3 Medical Testing

There is an unpleasant condition called *BO* suffered by 10% of the population. There are no prior symptoms; victims just suddenly start to stink. Fortunately, there is a test for latent *BO* before things start to smell. The test is not perfect, however:

- If you have the condition, there is a 10% chance that the test will say you do not have it. These are called "false negatives."
- If you do not have the condition, there is a 30% chance that the test will say you do. These are "false positives."

Suppose a random person is tested for latent *BO*. If the test is positive, then what is the probability that the person has the condition?

#### Step 1: Find the Sample Space

The sample space is found with the tree diagram in Figure 17.14.

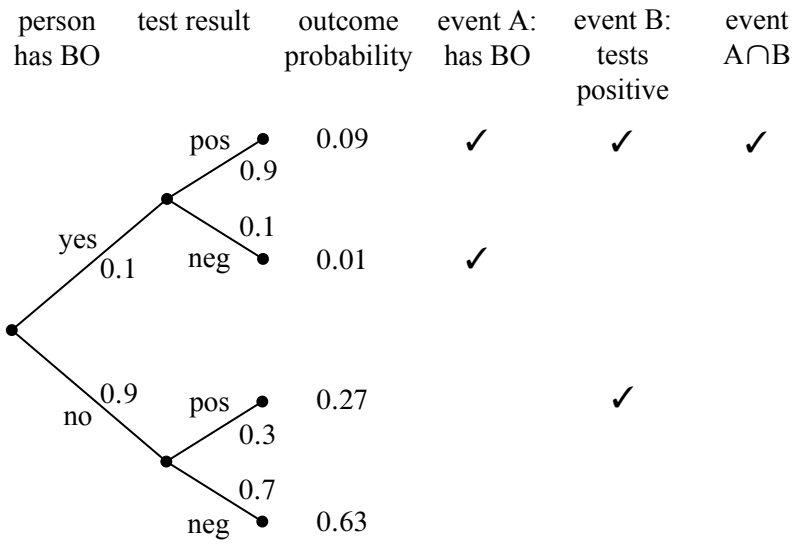



**Figure 17.14** The tree diagram for the BO problem.

**Step 2: Define Events of Interest**

Let $A$ be the event that the person has *BO*. Let $B$ be the event that the test was positive. The outcomes in each event are marked in the tree diagram. We want to find $\Pr[A \mid B]$, the probability that a person has *BO*, given that the test was positive.

**Step 3: Find Outcome Probabilities**

First, we assign probabilities to edges. These probabilities are drawn directly from the problem statement. By the Product Rule, the probability of an outcome is the product of the probabilities on the corresponding root-to-leaf path. All probabilities are shown in Figure 17.14.

**Step 4: Compute Event Probabilities**

From Definition 17.5.1, we have

$$\Pr[A \mid B] = \frac{\Pr[A \cap B]}{\Pr[B]} = \frac{0.09}{0.09 + 0.27} = \frac{1}{4}.$$

So, if you test positive, then there is only a 25% chance that you have the condition!

This answer is initially surprising, but makes sense on reflection. There are two ways you could test positive. First, it could be that you have the condition and the test is correct. Second, it could be that you are healthy and the test is incorrect. The

problem is that almost everyone is healthy; therefore, most of the positive results arise from incorrect tests of healthy people!

We can also compute the probability that the test is correct for a random person. This event consists of two outcomes. The person could have the condition and test positive (probability 0.09), or the person could be healthy and test negative (probability 0.63). Therefore, the test is correct with probability $0.09 + 0.63 = 0.72$. This is a relief; the test is correct almost three-quarters of the time.

But wait! There is a simple way to make the test correct 90% of the time: always return a negative result! This "test" gives the right answer for all healthy people and the wrong answer only for the 10% that actually have the condition. So a better strategy by this measure is to completely ignore the test result!

There is a similar paradox in weather forecasting. During winter, almost all days in Boston are wet and overcast. Predicting miserable weather every day may be more accurate than really trying to get it right!

### 17.5.4 *A Posteriori* Probabilities

If you think about it too much, the medical testing problem we just considered could start to trouble you. The concern would be that by the time you take the test, you either have the BO condition or you don't —you just don't know which it is. So you may wonder if a statement like "If you tested positive, then you have the condition with probability 25%" makes sense.

In fact, such a statement does make sense. It means that 25% of the people who test positive actually have the condition. It is true that any particular person has it or they don't, but a *randomly selected* person among those who test positive will have the condition with probability 25%.

Anyway, if the medical testing example bothers you, you will definitely be worried by the following examples, which go even further down this path.

### 17.5.5 The "Halting Problem," in Reverse

Suppose that we turn the hockey question around: what is the probability that the Halting Problem won their first game, given that they won the series?

This seems like an absurd question! After all, if the Halting Problem won the series, then the winner of the first game has already been determined. Therefore, who won the first game is a question of fact, not a question of probability. However, our mathematical theory of probability contains no notion of one event preceding another—there is no notion of time at all. Therefore, from a mathematical perspective, this is a perfectly valid question. And this is also a meaningful question from a practical perspective. Suppose that you're told that the Halting Problem won the series, but not told the results of individual games. Then, from your perspective, it

makes perfect sense to wonder how likely it is that The Halting Problem won the first game.

A conditional probability $\Pr\left[B \mid A\right]$ is called *a posteriori* if event $B$ precedes event $A$ in time. Here are some other examples of a posteriori probabilities:

- The probability it was cloudy this morning, given that it rained in the afternoon.
- The probability that I was initially dealt two queens in Texas No Limit Hold 'Em poker, given that I eventually got four-of-a-kind.

Mathematically, a posteriori probabilities are *no different* from ordinary probabilities; the distinction is only at a higher, philosophical level. Our only reason for drawing attention to them is to say, "Don't let them rattle you."

Let's return to the original problem. The probability that the Halting Problem won their first game, given that they won the series is $\Pr\left[B \mid A\right]$. We can compute this using the definition of conditional probability and the tree diagram in Figure 17.13:

$$\Pr\left[B \mid A\right] = \frac{\Pr[B \cap A]}{\Pr[A]} = \frac{1/3 + 1/18}{1/3 + 1/18 + 1/9} = \frac{7}{9}.$$

This answer is suspicious! In the preceding section, we showed that $\Pr\left[A \mid B\right]$ was also $7/9$. Could it be true that $\Pr\left[A \mid B\right] = \Pr\left[B \mid A\right]$ in general? Some reflection suggests this is unlikely. For example, the probability that I feel uneasy, given that I was abducted by aliens, is pretty large. But the probability that I was abducted by aliens, given that I feel uneasy, is rather small.

Let's work out the general conditions under which $\Pr\left[A \mid B\right] = \Pr\left[B \mid A\right]$. By the definition of conditional probability, this equation holds if an only if:

$$\frac{\Pr[A \cap B]}{\Pr[B]} = \frac{\Pr[A \cap B]}{\Pr[A]}$$

This equation, in turn, holds only if the denominators are equal or the numerator is 0; namely if

$$\Pr[B] = \Pr[A] \qquad \text{or} \qquad \Pr[A \cap B] = 0.$$

The former condition holds in the hockey example; the probability that the Halting Problem wins the series (event $A$) is equal to the probability that it wins the first game (event $B$) since both probabilities are $1/2$.

In general, such pairs of probabilities are related by Bayes' Rule:

**Theorem 17.5.2** (Bayes' Rule)**.** *If* $\Pr[A]$ *and* $\Pr[B]$ *are nonzero, then:*

$$\Pr\left[B \mid A\right] = \frac{\Pr\left[A \mid B\right] \cdot \Pr[B]}{\Pr[A]} \tag{17.5}$$

*Proof.* When $\Pr[A]$ and $\Pr[B]$ are nonzero, we have

$$\Pr\left[A \mid B\right] \cdot \Pr[B] = \Pr[A \cap B] = \Pr\left[B \mid A\right] \cdot \Pr[A]$$

by definition of conditional probability. Dividing by $\Pr[A]$ gives (17.5). ■

### 17.5.6 The Law of Total Probability

Breaking a probability calculation into cases simplifies many problems. The idea is to calculate the probability of an event $A$ by splitting into two cases based on whether or not another event $E$ occurs. That is, calculate the probability of $A \cap E$ and $A \cap \overline{E}$. By the Sum Rule, the sum of these probabilities equals $\Pr[A]$. Expressing the intersection probabilities as conditional probabilities yields:

**Rule 17.5.3** (Law of Total Probability, single event)**.** *If* $\Pr[E]$ *and* $\Pr[\overline{E}]$ *are nonzero, then*

$$\Pr[A] = \Pr\left[A \mid E\right] \cdot \Pr[E] + \Pr\left[A \mid \overline{E}\right] \cdot \Pr[\overline{E}].$$

For example, suppose we conduct the following experiment. First, we flip a fair coin. If heads comes up, then we roll one die and take the result. If tails comes up, then we roll two dice and take the sum of the two results. What is the probability that this process yields a 2? Let $E$ be the event that the coin comes up heads, and let $A$ be the event that we get a 2 overall. Assuming that the coin is fair, $\Pr[E] = \Pr[\overline{E}] = 1/2$. There are now two cases. If we flip heads, then we roll a 2 on a single die with probability $\Pr\left[A \mid E\right] = 1/6$. On the other hand, if we flip tails, then we get a sum of 2 on two dice with probability $\Pr\left[A \mid \overline{E}\right] = 1/36$. Therefore, the probability that the whole process yields a 2 is

$$\Pr[A] = \frac{1}{2} \cdot \frac{1}{6} + \frac{1}{2} \cdot \frac{1}{36} = \frac{7}{72}.$$

There is also a form of the rule to handle more than two cases.

**Rule 17.5.4** (Law of Total Probability)**.** *If* $E_1, \ldots, E_n$ *are disjoint events whose union is the whole sample space, then:*

$$\Pr[A] = \sum_{i=1}^{n} \Pr\left[A \mid E_i\right] \cdot \Pr[E_i].$$

### 17.5.7 Conditioning on a Single Event

The probability rules that we derived in Section 17.4.2 extend to probabilities conditioned on the same event. For example, the Inclusion-Exclusion formula for two sets holds when all probabilities are conditioned on an event $C$:

$$\Pr[A \cup B \mid C] = \Pr[A \mid C] + \Pr[B \mid C] - \Pr[A \cap B \mid C].$$

This is easy to verify by plugging in the Definition 17.5.1 of conditional probability.[5]

It is important not to mix up events before and after the conditioning bar. For example, the following is *not* a valid identity:

**False Claim.**

$$\Pr[A \mid B \cup C] = \Pr[A \mid B] + \Pr[A \mid C] - \Pr[A \mid B \cap C]. \tag{17.6}$$

A simple counter-example is to let $B$ and $C$ be events over a uniform space with most of their outcomes in $A$, but not overlapping. This ensures that $\Pr[A \mid B]$ and $\Pr[A \mid C]$ are both close to 1. For example,

$$\begin{aligned} B &::= [0, 9], \\ C &::= [10, 18] \cup \{0\}, \\ A &::= [1, 18], \end{aligned}$$

so

$$\Pr[A \mid B] = \frac{9}{10} = \Pr[A \mid C].$$

Also, since 0 is the only outcome in $B \cap C$ and $0 \notin A$, we have

$$\Pr[A \mid B \cap C] = 0$$

So the right hand side of (17.6) is 1.8, while the left hand side is a probability which can be at most 1 —actually, it is 18/19.

### 17.5.8 Discrimination Lawsuit

Several years ago there was a sex discrimination lawsuit against a famous university. A woman math professor was denied tenure, allegedly because she was a woman. She argued that in every one of the university's 22 departments, the percentage of men candidates granted tenure was greater than the percentage of women candidates granted tenure. This sounds very suspicious!

[5] Problem 17.11 explains why this and similar conditional identities follow on general principles from the corresponding unconditional identities.

However, the university's lawyers argued that across the university as a whole, the percentage of male candidates granted tenure was actually *lower* than the percentage for women candidates. This suggests that if there was any sex discrimination, then it was against men! Surely, at least one party in the dispute must be lying.

Let's clarify the problem by expressing both arguments in terms of conditional probabilities. To simplify matters, suppose that there are only two departments, EE and CS, and consider the experiment where we pick a random candidate. Define the following events:

- $A ::=$ the candidate is granted tenure,
- $F_{EE} ::=$ the candidate is a woman in the EE department,
- $F_{CS} ::=$ the candidate is a woman in the CS department,
- $M_{EE} ::=$ the candidate is a man in the EE department,
- $M_{CS} ::=$ the candidate is a man in the CS department.

Assume that all candidates are either men or women, and that no candidate belongs to both departments. That is, the events $F_{EE}$, $F_{CS}$, $M_{EE}$, and $M_{CS}$ are all disjoint.

In these terms, the plaintiff is making the following argument:

$$\Pr\left[A \mid F_{EE}\right] < \Pr\left[A \mid M_{EE}\right] \quad \text{and}$$
$$\Pr\left[A \mid F_{CS}\right] < \Pr\left[A \mid M_{CS}\right].$$

That is, in both departments, the probability that a woman candidate is granted tenure is less than the probability for a man.

The university retorts that *overall*, a woman candidate is *more* likely to be granted tenure than a man; namely that

$$\Pr\left[A \mid F_{EE} \cup F_{CS}\right] > \Pr\left[A \mid M_{EE} \cup M_{CS}\right].$$

It is easy to believe that these two positions are contradictory, and the phenomenon illustrated here is widely referred to as "Simpson's Paradox." But there is no contradiction or paradox, and in fact, Table 17.1 shows a set of candidate statistics for which the assertions of both the plaintiff and the university hold. In this case, a higher percentage of men candidates were granted tenure in each department, but overall a higher percentage of women candidates were granted tenure! How do we make sense of this?

| | | |
|---|---|---|
| CS | 0 women granted tenure, 1 candidate | 0% |
| | 50 men granted tenure, 100 candidate | 50% |
| EE | 70 women granted tenure, 100 candidate | 70% |
| | 1 man granted tenure, 1 candidate | 100% |
| Overall | 70 women granted tenure, 101 candidate | $\approx 70\%$ |
| | 51 men granted tenure, 101 candidate | $\approx 51\%$ |

**Table 17.1** A scenario where women are less likely to be granted tenure than men in each department, but more likely to be granted tenure overall.

With data like this showing that at the department level, women candidates were less likely to be granted tenure than men, university administrators would likely see an indication of bias against women, and the departments would be directed to reexamine their admission procedures.

But suppose we replaced "the candidate is a man/woman in the EE department," by "the candidate is a man/woman for whom a tenure decision was made during an odd-numbered day of the month," and likewise with CS and an even-numbered day of the month. Since we don't think the parity of a date is a cause for the outcome of a tenure decision, we would ignore the "coincidence" that on both odd and even dates, men are more frequently granted tenure. Instead, we would judge, based on the overall data showing women more likely to be granted tenure, that gender bias against women was *not* an issue in the university.

The point is that it's the *same data* that we interpret differently based on our implicit causal beliefs. It would be circular to claim that the gender correlation observed in the data corroborates our belief that there is discrimination, since our interpretation of the data correlation *depends* on our beliefs about the causes of tenure decisions.[6]. This illustrates a basic principle in statistics which people constantly ignore: *never assume that correlation implies causation*.

## 17.6 Independence

Suppose that we flip two fair coins simultaneously on opposite sides of a room. Intuitively, the way one coin lands does not affect the way the other coin lands. The mathematical concept that captures this intuition is called *independence*.

[6]These issues are thoughtfully examined in *Causality: Models, Reasoning and Inference*, Judea Pearl, Cambridge U. Press, 2001

**Definition 17.6.1.** An event with probability 0 is defined to be independent of every event (including itself). If $\Pr[B] \neq 0$, then event $A$ is independent of event $B$ iff

$$\Pr\left[A \mid B\right] = \Pr[A]. \tag{17.7}$$

In other words, $A$ and $B$ are independent if knowing that $B$ happens does not alter the probability that $A$ happens, as is the case with flipping two coins on opposite sides of a room.

#### Potential Pitfall

Students sometimes get the idea that disjoint events are independent. The *opposite* is true: if $A \cap B = \emptyset$, then knowing that $A$ happens means you know that $B$ does not happen. So disjoint events are *never* independent—unless one of them has probability zero.

### 17.6.1 Alternative Formulation

Sometimes it is useful to express independence in an alternate form which follows immediately from Definition 17.6.1:

**Theorem 17.6.2.** *$A$ is independent of $B$ if and only if*

$$\Pr[A \cap B] = \Pr[A] \cdot \Pr[B]. \tag{17.8}$$

Notice that Theorem 17.6.2 makes apparent the symmetry between $A$ being independent of $B$ and $B$ being independent of $A$:

**Corollary 17.6.3.** *$A$ is independent of $B$ iff $B$ is independent of $A$.*

### 17.6.2 Independence Is an Assumption

Generally, independence is something that you *assume* in modeling a phenomenon. For example, consider the experiment of flipping two fair coins. Let $A$ be the event that the first coin comes up heads, and let $B$ be the event that the second coin is heads. If we assume that $A$ and $B$ are independent, then the probability that both coins come up heads is:

$$\Pr[A \cap B] = \Pr[A] \cdot \Pr[B] = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}.$$

In this example, the assumption of independence is reasonable. The result of one coin toss should have negligible impact on the outcome of the other coin toss. And if we were to repeat the experiment many times, we would be likely to have $A \cap B$ about 1/4 of the time.

There are, of course, many examples of events where assuming independence is *not* justified, For example, let $C$ be the event that tomorrow is cloudy and $R$ be the event that tomorrow is rainy. Perhaps $\Pr[C] = 1/5$ and $\Pr[R] = 1/10$ in Boston. If these events were independent, then we could conclude that the probability of a rainy, cloudy day was quite small:

$$\Pr[R \cap C] = \Pr[R] \cdot \Pr[C] = \frac{1}{5} \cdot \frac{1}{10} = \frac{1}{50}.$$

Unfortunately, these events are definitely not independent; in particular, every rainy day is cloudy. Thus, the probability of a rainy, cloudy day is actually $1/10$.

Deciding when to *assume* that events are independent is a tricky business. In practice, there are strong motivations to assume independence since many useful formulas (such as equation (17.8)) only hold if the events are independent. But you need to be careful: we'll describe several famous examples where (false) assumptions of independence led to trouble. This problem gets even trickier when there are more than two events in play.

### 17.6.3 Mutual Independence

We have defined what it means for two events to be independent. What if there are more than two events? For example, how can we say that the flips of $n$ coins are all independent of one another? A set of events is said to be *mutually independent* if the probability of each event in the set is the same no matter which of the other events has occurred. We could formalize this with conditional probabilities as in Definition 17.6.1, but we'll jump directly to the cleaner definition based on products of probabilities as in Theorem 17.6.2:

**Definition 17.6.4.** A set of events $E_1, E_2, \ldots, E_n$ is mutually independent iff for all subsets $S \subseteq [1, n]$,

$$\Pr\left[\bigcap_{j \in S} E_j\right] = \prod_{j \in S} \Pr[E_j].$$

Definition 17.6.4 says that $E_1, E_2, \ldots, E_n$ are mutually independent if and only if all of the following equations hold for all distinct $i$, $j$, $k$, and $l$:

$$\begin{aligned}
\Pr[E_i \cap E_j] &= \Pr[E_i] \cdot \Pr[E_j] \\
\Pr[E_i \cap E_j \cap E_k] &= \Pr[E_i] \cdot \Pr[E_j] \cdot \Pr[E_k] \\
\Pr[E_i \cap E_j \cap E_k \cap E_l] &= \Pr[E_i] \cdot \Pr[E_j] \cdot \Pr[E_k] \cdot \Pr[E_l] \\
&\vdots \\
\Pr[E_1 \cap \cdots \cap E_n] &= \Pr[E_1] \cdots \Pr[E_n].
\end{aligned}$$

For example, if we toss $n$ fair coins, the tosses are mutually independent iff for every subset of $m$ coins, the probability that every coin in the subset comes up heads is $2^{-m}$.

### 17.6.4 DNA Testing

Assumptions about independence are routinely made in practice. Frequently, such assumptions are quite reasonable. Sometimes, however, the reasonableness of an independence assumption is not so clear, and the consequences of a faulty assumption can be severe.

For example, consider the following testimony from the O. J. Simpson murder trial on May 15, 1995:

**Mr. Clarke:** When you make these estimations of frequency—and I believe you touched a little bit on a concept called independence?

**Dr. Cotton:** Yes, I did.

**Mr. Clarke:** And what is that again?

**Dr. Cotton:** It means whether or not you inherit one allele that you have is not—does not affect the second allele that you might get. That is, if you inherit a band at 5,000 base pairs, that doesn't mean you'll automatically or with some probability inherit one at 6,000. What you inherit from one parent is what you inherit from the other.

**Mr. Clarke:** Why is that important?

**Dr. Cotton:** Mathematically that's important because if that were not the case, it would be improper to multiply the frequencies between the different genetic locations.

**Mr. Clarke:** How do you—well, first of all, are these markers independent that you've described in your testing in this case?

Presumably, this dialogue was as confusing to you as it was for the jury. Essentially, the jury was told that genetic markers in blood found at the crime scene matched Simpson's. Furthermore, they were told that the probability that the markers would be found in a randomly-selected person was at most 1 in 170 million. This astronomical figure was derived from statistics such as:

- 1 person in 100 has marker $A$.
- 1 person in 50 marker $B$.

- 1 person in 40 has marker $C$.
- 1 person in 5 has marker $D$.
- 1 person in 170 has marker $E$.

Then these numbers were multiplied to give the probability that a randomly-selected person would have all five markers:

$$\begin{aligned}\Pr[A \cap B \cap C \cap D \cap E] &= \Pr[A] \cdot \Pr[B] \cdot \Pr[C] \cdot \Pr[D] \cdot \Pr[E] \\ &= \frac{1}{100} \cdot \frac{1}{50} \cdot \frac{1}{40} \cdot \frac{1}{5} \cdot \frac{1}{170} = \frac{1}{170{,}000{,}000}.\end{aligned}$$

The defense pointed out that this assumes that the markers appear mutually independently. Furthermore, all the statistics were based on just a few hundred blood samples.

After the trial, the jury was widely mocked for failing to "understand" the DNA evidence. If you were a juror, would *you* accept the 1 in 170 million calculation?

### 17.6.5 Pairwise Independence

The definition of mutual independence seems awfully complicated—there are so many subsets of events to consider! Here's an example that illustrates the subtlety of independence when more than two events are involved. Suppose that we flip three fair, mutually-independent coins. Define the following events:

- $A_1$ is the event that coin 1 matches coin 2.
- $A_2$ is the event that coin 2 matches coin 3.
- $A_3$ is the event that coin 3 matches coin 1.

Are $A_1$, $A_2$, $A_3$ mutually independent?

The sample space for this experiment is:

$$\{HHH,\ HHT,\ HTH,\ HTT,\ THH,\ THT,\ TTH,\ TTT\}.$$

Every outcome has probability $(1/2)^3 = 1/8$ by our assumption that the coins are mutually independent.

To see if events $A_1$, $A_2$, and $A_3$ are mutually independent, we must check a sequence of equalities. It will be helpful first to compute the probability of each event $A_i$:

$$\begin{aligned}\Pr[A_1] &= \Pr[HHH] + \Pr[HHT] + \Pr[TTH] + \Pr[TTT] \\ &= \frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8} = \frac{1}{2}.\end{aligned}$$

By symmetry, $\Pr[A_2] = \Pr[A_3] = 1/2$ as well. Now we can begin checking all the equalities required for mutual independence in Definition 17.6.4:

$$\begin{aligned}\Pr[A_1 \cap A_2] &= \Pr[HHH] + \Pr[TTT] = \frac{1}{8} + \frac{1}{8} = \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2} \\ &= \Pr[A_1]\Pr[A_2].\end{aligned}$$

By symmetry, $\Pr[A_1 \cap A_3] = \Pr[A_1] \cdot \Pr[A_3]$ and $\Pr[A_2 \cap A_3] = \Pr[A_2] \cdot \Pr[A_3]$ must hold also. Finally, we must check one last condition:

$$\begin{aligned}\Pr[A_1 \cap A_2 \cap A_3] &= \Pr[HHH] + \Pr[TTT] = \frac{1}{8} + \frac{1}{8} = \frac{1}{4} \\ &\neq \frac{1}{8} = \Pr[A_1]\Pr[A_2]\Pr[A_3].\end{aligned}$$

The three events $A_1$, $A_2$, and $A_3$ are not mutually independent even though any two of them are independent! This not-quite mutual independence seems weird at first, but it happens. It even generalizes:

**Definition 17.6.5.** A set $A_1, A_2, \ldots,$ of events is *$k$-way independent* iff every set of $k$ of these events is mutually independent. The set is *pairwise independent* iff it is 2-way independent.

So the sets $A_1$, $A_2$, $A_3$ above are pairwise independent, but not mutually independent. Pairwise independence is a much weaker property than mutual independence.

For example, suppose that the prosecutors in the O. J. Simpson trial were wrong and markers $A$, $B$, $C$, $D$, and $E$ appear only *pairwise* independently. Then the probability that a randomly-selected person has all five markers is no more than:

$$\begin{aligned}\Pr[A \cap B \cap C \cap D \cap E] &\leq \Pr[A \cap E] = \Pr[A] \cdot \Pr[E] \\ &= \frac{1}{100} \cdot \frac{1}{170} = \frac{1}{17{,}000}.\end{aligned}$$

The first line uses the fact that $A \cap B \cap C \cap D \cap E$ is a subset of $A \cap E$. (We picked out the $A$ and $E$ markers because they're the rarest.) We use pairwise independence on the second line. Now the probability of a random match is 1 in 17,000 —a far cry from 1 in 170 million! And this is the strongest conclusion we can reach assuming only pairwise independence.

On the other hand, the 1 in 17,000 bound that we get by assuming pairwise independence is a lot better than the bound that we would have if there were no independence at all. For example, if the markers are dependent, then it is possible that

everyone with marker $E$ has marker $A$,

everyone with marker $A$ has marker $B$,

everyone with marker $B$ has marker $C$, and

everyone with marker $C$ has marker $D$.

In such a scenario, the probability of a match is

$$\Pr[E] = \frac{1}{170}.$$

So a stronger independence assumption leads to a smaller bound on the probability of a match. The trick is to figure out what independence assumption is reasonable. Assuming that the markers are *mutually* independent may well *not* be reasonable unless you have examined hundreds of millions of blood samples. Otherwise, how would you know that marker $D$ does not show up more frequently whenever the other four markers are simultaneously present?

We will conclude our discussion of independence with a useful, and somewhat famous, example known as the Birthday Principle.

### 17.6.6 The Birthday Principle

There are 95 students in a class. What is the probability that some birthday is shared by two people? Comparing 95 students to the 365 possible birthdays, you might guess the probability lies somewhere around $1/4$ —but you'd be wrong: the probability that there will be two people in the class with matching birthdays is actually more than 0.9999.

To work this out, we'll assume that the probability that a randomly chosen student has a given birthday is $1/d$, where $d = 365$ in this case. We'll also assume that a class is composed of $n$ randomly and independently selected students, with $n = 95$ in this case. These randomness assumptions are not really true, since more babies are born at certain times of year, and students' class selections are typically not independent of each other, but simplifying in this way gives us a start on analyzing the problem. More importantly, these assumptions are justifiable in important computer science applications of birthday matching. For example, the birthday matching is a good model for collisions between items randomly inserted into a hash table. So we won't worry about things like Spring procreation preferences that make January birthdays more common, or about twins' preferences to take classes together (or not).

Selecting a sequence of $n$ students for a class yields a sequence of $n$ birthdays. Under the assumptions above, the $d^n$ possible birthday sequences are equally likely outcomes. Let's examine the consequences of this probability model by focussing

on the $i$th and $j$th elements in a birthday sequence, where $1 \leq i \neq j \leq n$. It makes for a better story if we refer to the $i$th birthday as "Alice's" and the $j$th as "Bob's."

Now if Alice, Bob, Carol, and Don are four different people, then whether Alice and Bob have matching birthdays is independent of whether Carol and Don do. What's more interesting is that whether Alice and *Carol* have the same birthday is independent of whether Alice and Bob do. This follows because Carol is as likely to have the same birthday as Alice, independently of whatever birthdays Alice and Bob happen to have; a formal proof of this claim appears in Problem 18.2. In short, the set of all events that a couple has matching birthdays is *pairwise* independent, even for overlapping couples. This will be important Chapter 19 because pairwise independence will be enough to justify some conclusions about the expected number of matches. However, these matching birthday events are obviously *not* even 3-way independent: if Alice and Bob match, and also Alice and Carol match, then Bob and Carol will match.

It turns out that as long as the number of students is noticeably smaller than the number of possible birthdays, we can get a pretty good estimate of the birthday matching probabilities by *pretending* that the matching events are mutually independent. (An intuitive justification for this is that with only a small number of matching pairs, it's likely that none of the pairs overlap.) Then the probability of *no* matching birthdays would be the same as $r$th power of the probability that a couple does *not* have matching birthdays, where $r ::= \binom{n}{2}$ is the number of couples. That is, the probability of no matching birthdays would be

$$(1 - 1/d)^{\binom{n}{2}}. \tag{17.9}$$

Using the fact that $1 + x < e^x$ for all $x$,[7] we would conclude that the probability of no matching birthdays is at most

$$e^{-\binom{n}{2}/d}. \tag{17.10}$$

The matching birthday problem fits in here so far as a nice example illustrating pairwise and mutual independence, but it's actually not hard to justify the bound (17.10) without any pretence of independence. Namely, there are $d(d-1)(d-2)\cdots(d-(n-1))$ length $n$ sequences of distinct birthdays. So the proba-

[7] This approximation is obtained by truncating the Taylor series $e^{-x} = 1 - x + x^2/2! - x^3/3! + \cdots$. The approximation $e^{-x} \approx 1 - x$ is pretty accurate when $x$ is small.

bility that everyone has a different birthday is:

$$\begin{aligned}
&\frac{d(d-1)(d-2)\cdots(d-(n-1))}{d^n} \\
&= \frac{d}{d}\cdot\frac{d-1}{d}\cdot\frac{d-2}{d}\cdots\frac{d-(n-1)}{d} \\
&= \left(1-\frac{0}{d}\right)\left(1-\frac{1}{d}\right)\left(1-\frac{2}{d}\right)\cdots\left(1-\frac{n-1}{d}\right) \\
&< e^0\cdot e^{-1/d}\cdot e^{-2/d}\cdots e^{-(n-1)/d} && \text{(since } 1+x<e^x\text{)} \\
&= e^{-\left(\sum_{i=1}^{n-1} i/d\right)} \\
&= e^{-(n(n-1)/2d)} \\
&= \text{the bound (17.10).}
\end{aligned}$$

For $n = 85$ and $d = 365$, the value of (17.10) is less than $1/17,000$, which means the probability of having some pair of matching birthdays actually is more than $1 - 1/17,000 > 0.9999$. So it would be pretty astonishing if there were no pair of students in the class with matching birthdays.

For $d \leq n^2/2$, the probability of no match turns out to be asymptotically equal to the upper bound (17.10). For $d = n^2/2$ in particular, the probability of no match is asymptotically equal to $1/e$. This leads to a rule of thumb which is useful in many contexts in computer science:

> **The Birthday Principle**
>
> If there are $d$ days in a year and $\sqrt{2d}$ people in a room, then the probability that two share a birthday is about $1 - 1/e \approx 0.632$.

For example, the Birthday Principle says that if you have $\sqrt{2\cdot 365} \approx 27$ people in a room, then the probability that two share a birthday is about 0.632. The actual probability is about 0.626, so the approximation is quite good.

Among other applications, it implies that to use a hash function that maps $n$ items into a hash table of size $d$, you can expect many collisions unless $n^2$ is a small fraction of $d$. The Birthday Principle also famously comes into play as the basis of "birthday attacks" that crack certain cryptographic systems.

## Problems for Section 17.2

### Practice Problems

**Problem 17.1.**
Let $B$ be the number of heads that come up on $2n$ independent tosses of a fair coin.
**(a)** $\Pr[B = n]$ is asymptotically equal to one of the expressions given below. Explain which one.

1. $\frac{1}{\sqrt{2\pi n}}$
2. $\frac{2}{\sqrt{\pi n}}$
3. $\frac{1}{\sqrt{\pi n}}$
4. $\sqrt{\frac{2}{\pi n}}$

**Problem 17.2.**
Suppose you flip a fair coin 100 times. The coin flips are all mutually independent.

**(a)** What is the expected number of heads?

**(b)** What upper bound on the probability that the number of heads is at least 70 can we derive using Markov's Theorem?

**(c)** What is the variance of the number of heads?

**(d)** What upper bound does Chebyshev's Theorem give us on the probability that the number of heads is either less than 30 or greater than 70?

### Exam Problems

**Problem 17.3.** **(a)** What's the probability that 0 doesn't appear among $k$ digits chosen independently and uniformly at random?

**(b)** A box contains 90 good and 10 defective screws. What's the probability that if we pick 10 screws from the box, none will be defective?

**(c)** First one digit is chosen uniformly at random from $\{1, 2, 3, 4, 5\}$ and is removed from the set; then a second digit is chosen uniformly at random from the remaining digits. What is the probability that an odd digit is picked the second time?

**(d)** Suppose that you *randomly* permute the digits $1, 2, \cdots, n$, that is, you select a permutation uniformly at random. What is the probability the digit $k$ ends up in the $i$th position after the permutation?

**(e)** A fair coin is flipped $n$ times. What's the probability that all the heads occur at the end of the sequence? (If no heads occur, then "all the heads are at the end of the sequence" is vacuously true.)

### Class Problems

**Problem 17.4.**
In the alternate universe where the Red Sox don't regularly collapse at the end of their season, the New York Yankees and the Boston Red Sox are playing a two-out-of-three series. (In other words, they play until one team has won two games. Then that team is declared the overall winner and the series ends. Again, a fantasy.) Assume that the Red Sox win each game with probability $3/5$, regardless of the outcomes of previous games.

Answer the questions below using the four step method. You can use the same tree diagram for all three problems.

**(a)** What is the probability that a total of 3 games are played?

**(b)** What is the probability that the winner of the series loses the first game?

**(c)** What is the probability that the *correct* team wins the series?

**Problem 17.5.**
To determine which of two people gets a prize, a coin is flipped twice. If the flips are a Head and then a Tail, the first player wins. If the flips are a Tail and then a Head, the second player wins. However, if both coins land the same way, the flips don't count and whole the process starts over.

Assume that on each flip, a Head comes up with probability $p$, regardless of what happened on other flips. Use the four step method to find a simple formula for the probability that the first player wins. What is the probability that neither player wins?

Suggestions: The tree diagram and sample space are infinite, so you're not going to finish drawing the tree. Try drawing only enough to see a pattern. Summing all the winning outcome probabilities directly is difficult. However, a neat trick solves this problem and many others. Let $s$ be the sum of all winning outcome probabilities in the whole tree. Notice that *you can write the sum of all the winning*

*probabilities in certain subtrees as a function of $s$*. Use this observation to write an equation in $s$ and then solve.

**Problem 17.6.**

Suppose you need a fair coin to decide which door to choose in the 6.042 Monty Hall game. After making everyone in your group empty their pockets, all you managed to turn up is some old collaboration statements, a few used tissues, and one penny. However, the penny was from Prof. Meyer's pocket, so it is **not** safe to assume that it is a fair coin.

How can we use a coin of unknown bias to get the same effect as a fair coin of bias $1/2$? Draw the tree diagram for your solution, but since it is infinite, draw only enough to see a pattern.

Suggestion: A neat trick allows you to sum all the outcome probabilities that cause you to say "Heads": Let $s$ be the sum of all "Heads" outcome probabilities in the whole tree. Notice that *you can write the sum of all the "Heads" outcome probabilities in certain subtrees as a function of $s$*. Use this observation to write an equation in $s$ and then solve.

## Homework Problems

**Problem 17.7.**

Let's see what happens when *Let's Make a Deal* is played with **four** doors. A prize is hidden behind one of the four doors. Then the contestant picks a door. Next, the host opens an unpicked door that has no prize behind it. The contestant is allowed to stick with their original door or to switch to one of the two unopened, unpicked doors. The contestant wins if their final choice is the door hiding the prize.

Let's make the same assumptions as in the original problem:

1. The prize is equally likely to be behind each door.
2. The contestant is equally likely to pick each door initially, regardless of the prize's location.
3. The host is equally likely to reveal each door that does not conceal the prize and was not selected by the player.

Use The Four Step Method to find the following probabilities. The tree diagram may become awkwardly large, in which case just draw enough of it to make its structure clear.

**(a)** Contestant Stu, a sanitation engineer from Trenton, New Jersey, stays with his original door. What is the probability that Stu wins the prize?

**(b)** Contestant Zelda, an alien abduction researcher from Helena, Montana, switches to one of the remaining two doors with equal probability. What is the probability that Zelda wins the prize?

Now let's revise our assumptions about how contestants choose doors. Say the doors are labeled A, B, C, and D. Suppose that Carol always opens the *earliest* door possible (the door whose label is earliest in the alphabet) with the restriction that she can neither reveal the prize nor open the door that the player picked.

This gives contestant Mergatroid —an engineering student from Cambridge, MA —just a little more information about the location of the prize. Suppose that Mergatroid always switches to the earliest door, excluding his initial pick and the one Carol opened.

**(c)** What is the probability that Mergatroid wins the prize?

**Problem 17.8.**

I have a deck of 52 regular playing cards, 26 red, 26 black, randomly shuffled. They all lie face down in the deck so that you can't see them. I will draw a card off the top of the deck and turn it face up so that you can see it and then put it aside. I will continue to turn up cards like this but at some point while there are still cards left in the deck, you have to declare that you want the next card in the deck to be turned up. If that next card turns up black you win and otherwise you lose. Either way, the game is then over.

**(a)** Show that if you take the first card before you have seen any cards, you then have probability $1/2$ of winning the game.

**(b)** Suppose you don't take the first card and it turns up red. Show that you have then have a probability of winning the game that is greater than $1/2$.

**(c)** If there are $r$ red cards left in the deck and $b$ black cards, show that the probability of winning if you take the next card is $b/(r+b)$.

**(d)** Either,

1. come up with a strategy for this game that gives you a probability of winning strictly greater than $1/2$ and prove that the strategy works, or,
2. come up with a proof that no such strategy can exist.

## Problems for Section 17.4

### Class Problems

**Problem 17.9.**
Suppose there is a system, built by Caltech graduates, with $n$ components. We know from past experience that any particular component will fail in a given year with probability $p$. That is, letting $F_i$ be the event that the $i$th component fails within one year, we have

$$\Pr[F_i] = p$$

for $1 \leq i \leq n$. The *system* will fail if *any one* of its components fails. What can we say about the probability that the system will fail within one year?

Let $F$ be the event that the system fails within one year. Without any additional assumptions, we can't get an exact answer for $\Pr[F]$. However, we can give useful upper and lower bounds, namely,

$$p \leq \Pr[F] \leq np. \tag{17.11}$$

We may as well assume $p < 1/n$, since the upper bound is trivial otherwise. For example, if $n = 100$ and $p = 10^{-5}$, we conclude that there is at most one chance in 1000 of system failure within a year and at least one chance in 100,000.

Let's model this situation with the sample space $\mathcal{S} ::= \mathcal{P}([1, n])$ whose outcomes are subsets of positive integers $\leq n$, where $s \in \mathcal{S}$ corresponds to the indices of exactly those components that fail within one year. For example, $\{2, 5\}$ is the outcome that the second and fifth components failed within a year and none of the other components failed. So the outcome that the system did not fail corresponds to the emptyset, $\emptyset$.

**(a)** Show that the probability that the system fails could be as small as $p$ by describing appropriate probabilities for the outcomes. Make sure to verify that the sum of your outcome probabilities is 1.

**(b)** Show that the probability that the system fails could actually be as large as $np$ by describing appropriate probabilities for the outcomes. Make sure to verify that the sum of your outcome probabilities is 1.

**(c)** Prove inequality (17.11).

**Problem 17.10.**
Here are some handy rules for reasoning about probabilities that all follow directly from the Disjoint Sum Rule. Prove them.

$$\Pr[A - B] = \Pr[A] - \Pr[A \cap B] \qquad \text{(Difference Rule)}$$

$$\Pr[\overline{A}] = 1 - \Pr[A] \qquad \text{(Complement Rule)}$$

$$\Pr[A \cup B] = \Pr[A] + \Pr[B] - \Pr[A \cap B] \qquad \text{(Inclusion-Exclusion)}$$

$$\Pr[A \cup B] \leq \Pr[A] + \Pr[B] \qquad \text{(2-event Union Bound)}$$

$$\text{If } A \subseteq B, \text{ then } \Pr[A] \leq \Pr[B] \qquad \text{(Monotonicity)}$$

**Problem 17.11.**
Suppose $\Pr[] : \mathcal{S} \to [0, 1]$ is a probability function on a sample space, $\mathcal{S}$, and let $B$ be an event such that $\Pr[B] > 0$. Define a function $\Pr_B[\cdot]$ on outcomes $w \in \mathcal{S}$ by the rule:

$$\Pr_B[\omega] ::= \begin{cases} \Pr[\omega]/\Pr[B] & \text{if } \omega \in B, \\ 0 & \text{if } \omega \notin B. \end{cases} \tag{17.12}$$

**(a)** Prove that $\Pr_B[\cdot]$ is also a probability function on $\mathcal{S}$ according to Definition 17.4.2.

**(b)** Prove that

$$\Pr_B[A] = \frac{\Pr[A \cap B]}{\Pr[B]}$$

for all $A \subseteq \mathcal{S}$.

## Homework Problems

**Problem 17.12.**
Prove the following probabilistic identity, referred to as the **Union Bound**. You may assume the theorem that the probability of a union of *disjoint* sets is the sum of their probabilities.

Let $A_1, \ldots, A_n$ be a collection of events. Then

$$\Pr[A_1 \cup A_2 \cup \cdots \cup A_n] \leq \sum_{i=1}^{n} \Pr[A_i].$$

*Hint:* Induction.

## Problems for Section 17.5

### Practice Problems

**Problem 17.13.**
Dirty Harry places two bullets in the six-shell cylinder of his revolver. He gives the cylinder a random spin and says "Feeling lucky?" as he holds the gun against your heart.

**(a)** What is the probability that you will get shot if he pulls the trigger?

**(b)** Suppose he pulls the trigger and you don't get shot. What is the probability that you will get shot if he pulls the trigger a second time?

**(c)** Suppose you noticed that he placed the two shells next to each other in the cylinder. How does this change the answers to the previous two questions?

### Class Problems

**Problem 17.14.**
There are two decks of cards. One is complete, but the other is missing the Ace of spades. Suppose you pick one of the two decks with equal probability and then select a card from that deck uniformly at random. What is the probability that you picked the complete deck, given that you selected the eight of hearts? Use the four-step method and a tree diagram.

**Problem 17.15.**
Suppose you have three cards: A♡, A♠, and a Jack. From these, you choose a random hand (that is, each card is equally likely to be chosen) of two cards, and let $K$ be the number of Aces in your hand. You then randomly pick one of the cards in the hand and reveal it.

**(a)** Describe a simple probability space (that is, outcomes and their probabilities) for this scenario, and list the outcomes in each of the following events:

1. $[K \geq 1]$, (that is, your hand has an Ace in it),
2. A♡ is in your hand,
3. the revealed card is an A♡,
4. the revealed card is an Ace.

**(b)** Then calculate $\Pr\left[K = 2 \mid E\right]$ for $E$ equal to each of the four events in part (a). Notice that most, but *not all*, of these probabilities are equal.

Now suppose you have a deck with $d$ distinct cards, $a$ different kinds of Aces (including an A♡), you draw a random hand with $h$ cards, and then reveal a random card from your hand.

**(c)** Prove that Pr[A♡ is in your hand] $= h/d$.

**(d)** Prove that

$$\Pr\left[K = 2 \mid \text{A♡ is in your hand}\right] = \Pr[K = 2] \cdot \frac{2d}{ah}. \tag{17.13}$$

**(e)** Conclude that

$$\Pr\left[K = 2 \mid \text{the revealed card is an Ace}\right] = \Pr\left[K = 2 \mid \text{A♡ is in your hand}\right].$$

**Problem 17.16.**

There are three prisoners in a maximum-security prison for fictional villains: the Evil Wizard Voldemort, the Dark Lord Sauron, and Little Bunny Foo-Foo. The parole board has declared that it will release two of the three, chosen uniformly at random, but has not yet released their names. Naturally, Sauron figures that he will be released to his home in Mordor, where the shadows lie, with probability $2/3$.

A guard offers to tell Sauron the name of one of the other prisoners who will be released (either Voldemort or Foo-Foo). If the guard has a choice of naming either Voldemort or Foo-Foo (because both are to be released), he names one of the two with equal probability.

Sauron knows the guard to be a truthful fellow. However, Sauron declines this offer. He reasons that if the guard says, for example, "Little Bunny Foo-Foo will be released", then his own probability of release will drop to $1/2$. This is because he will then know that either he or Voldemort will also be released, and these two events are equally likely.

Dark Lord Sauron has made a typical mistake when reasoning about conditional probability. Using a tree diagram and the four-step method, explain his mistake. What is the probability that Sauron is released given that the guard says Foo-Foo is released?

*Hint:* Define the events $S$, $F$, and "$F$" as follows:

$$\begin{aligned} \text{"}F\text{"} &= \text{Guard says Foo-Foo is released} \\ F &= \text{Foo-Foo is released} \\ S &= \text{Sauron is released} \end{aligned}$$

## Homework Problems

**Problem 17.17.**
Outside of their hum-drum duties as Math for Computer Science Teaching Assistants, Oscar is trying to learn to levitate using only intense concentration and Liz is trying to become the world champion flaming torch juggler. Suppose that Oscar's probability of success is $1/6$, Liz's chance of success is $1/4$, and these two events are independent.

**(a)** If at least one of them succeeds, what is the probability that Oscar learns to levitate?

**(b)** If at most one of them succeeds, what is the probability that Liz becomes the world flaming torch juggler champion?

**(c)** If exactly one of them succeeds, what is the probability that it is Oscar?

**Problem 17.18.**
There is a course—not 6.042, naturally—in which 10% of the assigned problems contain errors. If you ask a Teaching Assistant (TA) whether a problem has an error, then they will answer correctly 80% of the time. This 80% accuracy holds regardless of whether or not a problem has an error. Likewise when you ask a lecturer, but with only 75% accuracy.

We formulate this as an experiment of choosing one problem randomly and asking a particular TA and Lecturer about it. Define the following events:

$$E ::= \text{"the problem has an error,"}$$
$$T ::= \text{"the TA says the problem has an error,"}$$
$$L ::= \text{"the lecturer says the problem has an error."}$$

**(a)** Translate the description above into a precise set of equations involving conditional probabilities among the events $E$, $T$, and $L$.

**(b)** Suppose you have doubts about a problem and ask a TA about it, and they tell you that the problem is correct. To double-check, you ask a lecturer, who says that the problem has an error. Assuming that *the correctness of the lecturers' answer and the TA's answer are independent of each other, regardless of whether there is an error*[8], what is the probability that there is an error in the problem?

[8]This assumption is questionable: by and large, we would expect the lecturer and the TA's to spot the same glaring errors and to be fooled by the same subtle ones.

**(c)** Is the event that "the TA says that there is an error", independent of the event that "the lecturer says that there is an error"?

**Problem 17.19.** **(a)** Suppose you repeatedly flip a fair coin until you see the sequence `HHT` or the sequence `TTH`. What is the probability you will see `HHT` first? *Hint:* Symmetry between Heads and Tails.

**(b)** What is the probability you see the sequence `HTT` before you see the sequence `HHT`? *Hint:* Try to find the probability that `HHT` comes before `HTT` conditioning on whether you first toss an `H` or a `T`. The answer is not $1/2$.

**Problem 17.20.**
A 52-card deck is thoroughly shuffled and you are dealt a hand of 13 cards.

**(a)** If you have one ace, what is the probability that you have a second ace?

**(b)** If you have the ace of spades, what is the probability that you have a second ace? Remarkably, the answer is different from part (a).

**Problem 17.21.**
You are organizing a neighborhood census and instruct your census takers to knock on doors and note the sex of any child that answers the knock. Assume that there are two children in a household and that girls and boys are equally likely to be children and to open the door.

A sample space for this experiment has outcomes that are triples whose first element is either `B` or `G` for the sex of the elder child, likewise for the second element and the sex of the younger child, and whose third coordinate is `E` or `Y` indicating whether the *e*lder child or *y*ounger child opened the door. For example, (`B`, `G`, `Y`) is the outcome that the elder child is a boy, the younger child is a girl, and the girl opened the door.

**(a)** Let $T$ be the event that the household has two girls, and $O$ be the event that a girl opened the door. List the outcomes in $T$ and $O$.

**(b)** What is the probability $\Pr[T \mid O]$, that both children are girls, given that a girl opened the door?

**(c)** Where is the mistake in the following argument?

If a girl opens the door, then we know that there is at least one girl in the household. The probability that there is at least one girl is

$$1 - \Pr[\text{both children are boys}] = 1 - (1/2 \times 1/2) = 3/4. \quad (17.14)$$

So,

$$\Pr\left[T \mid \text{there is at least one girl in the household}\right] \quad (17.15)$$

$$= \frac{\Pr[T \cap \text{there is at least one girl in the household}]}{\Pr[\text{there is at least one girl in the household}]} \quad (17.16)$$

$$= \frac{\Pr[T]}{\Pr[\text{there is at least one girl in the household}]} \quad (17.17)$$

$$= (1/4)/(3/4) = 1/3. \quad (17.18)$$

Therefore, given that a girl opened the door, the probability that there are two girls in the household is 1/3.

## Exam Problems

**Problem 17.22.**

Here's a variation of Monty Hall's game: the contestant still picks one of three doors, with a prize randomly placed behind one door and goats behind the other two. But now, instead of always opening a door to reveal a goat, Monty instructs Carol to *randomly* open one of the two doors that the contestant hasn't picked. This means she may reveal a goat, or she may reveal the prize. If she reveals the prize, then the entire game is *restarted*, that is, the prize is again randomly placed behind some door, the contestant again picks a door, and so on until Carol finally picks a door with a goat behind it. Then the contestant can choose to *stick* with his original choice of door or *switch* to the other unopened door. He wins if the prize is behind the door he finally chooses.

To analyze this setup, we define two events:

$GP$**:** The event that the contestant **g**uesses the door with the **p**rize behind it on his first guess.

$OP$**:** The event that the game is restarted at least once. Another way to describe this is as the event that the door Carol first **o**pens has a **p**rize behind it.

**(a)** What is $\Pr[GP]$? ... $\Pr\left[OP \mid \overline{GP}\right]$?

**(b)** What is $\Pr[OP]$?

**(c)** Let $R$ be the number of times the game is restarted before Carol picks a goat.

What is $\mathrm{Ex}[R]$? You may express the answer as a simple closed form in terms of $p ::= \Pr[OP]$.

**(d)** What is the probability the game will continue forever?

**(e)** When Carol finally picks the goat, the contestant has the choice of sticking or switching. Let's say that the contestant adopts the strategy of sticking. Let $W$ be the event that the contestant wins with this strategy, and let $w ::= \Pr[W]$. Express the following conditional probabilities as simple closed forms in terms of $w$.

i) $\Pr\left[W \mid GP\right] =$

ii) $\Pr\left[W \mid \overline{GP} \cap OP\right] =$

iii) $\Pr\left[W \mid \overline{GP} \cap \overline{OP}\right] =$

**(f)** What is $\Pr[W]$?

**(g)** For any final outcome where the contestant wins with a "stick" strategy, he would lose if he had used a "switch" strategy, and vice versa. In the original Monty Hall game, we concluded immediately that the probability that he would win with a "switch" strategy was $1 - \Pr[W]$. Why isn't this conclusion quite as obvious for this new, restartable game? Is this conclusion still sound? Briefly explain.

**Problem 17.23.**

There are two decks of cards, the red deck and the blue deck. They differ slightly in a way that makes drawing the eight of hearts slightly more likely from the red deck than from the blue deck.

One of the decks is randomly chosen and hidden in a box. You reach in the box and randomly pick a card that turns out to be the eight of hearts. You believe intuitively that this makes the red deck more likely to be in the box than the blue deck.

Your intuitive judgment about the red deck can be formalized and verified using some inequalities between probabilities and conditional probabilities involving the events

$R ::=$ **R**ed deck is in the box,

$B ::=$ **B**lue deck is in the box,

$E ::=$ **E**ight of hearts is picked from the deck in the box.

**(a)** State an inequality between probabilities and/or conditional probabilities that formalizes the assertion, "picking the eight of hearts from the red deck is more likely than from the blue deck."

**(b)** State a similar inequality that formalizes the assertion "picking the eight of hearts from the deck in the box makes the red deck more likely to be in the box than the blue deck."

**(c)** Assuming the each deck is equally likely to be the one in the box, prove that the inequality of part (a) implies the inequality of part (b).

**(d)** Suppose you couldn't be sure that the red deck and blue deck were equally likely to be in the box. Could you still conclude that picking the eight of hearts from the deck in the box makes the red deck more likely to be in the box than the blue deck? Briefly explain.

**Problem 17.24.**

There is a rare and serious disease called Beaver Fever which afflicts about 1 person in 1000. Victims of this disease start telling math jokes in social settings, believing other people will think they're funny.

Doctor Meyer has some fairly reliable tests for this disease. In particular:

- If a person has Beaver Fever, the probability that Meyer diagnoses the person as having the disease is 0.99.
- If a person doesn't have it, the probability that Meyer diagnoses that person as not having Beaver Fever is 0.97.

Let $B$ be the event that a randomly chosen person has Beaver Fever, and $Y$ be the event that Meyer's diagnosis is "Yes, that person has Beaver Fever," with $\overline{B}$ and $\overline{Y}$ the complements of these events.

**(a)** The description above explicitly gives the values of the following quantities. What are their values?

$$\Pr[B] \qquad \Pr\left[Y \mid B\right] \qquad \Pr\left[\overline{Y} \mid \overline{B}\right]$$

**(b)** Write formulas for $\Pr[\overline{B}]$ and $\Pr\left[Y \mid \overline{B}\right]$ solely in terms of the explicitly given expressions. Literally use the expressions, not their numeric values.

**(c)** Write a formula for the probability that Doctor Meyer says a person has the disease solely in terms of $\Pr[B]$, $\Pr[\overline{B}]$, $\Pr\left[Y \mid B\right]$ and $\Pr\left[Y \mid \overline{B}\right]$.

**(d)** Write a formula solely in terms of the expressions given in part (a) for the probability that a person has Beaver Fever given that Doctor Meyer says the person has it.

**Problem 17.25.**

Suppose that *Let's Make a Deal* is played according to slightly different rules and with a red goat and a blue goat. There are three doors, with a prize hidden behind one of them and the goats behind the others. No doors are opened until the contestant makes a final choice to stick or switch. The contestant is allowed to pick a door and ask a certain question that the host then answers honestly. The contestant may then stick with their chosen door, or switch to either of the other doors.

**(a)** If the contestant asks "is there is a goat behind one of the unchosen doors?" and the host answers "yes," is the contestant more likely to win the prize if they stick, switch, or does it not matter? Clearly identify the probability space of outcomes and their probabilities you use to model this situation. What is the contestant's probability of winning if he uses the best strategy?

**(b)** If the contestant asks "is the *red* goat behind one of the unchosen doors?" and the host answers "yes," is the contestant more likely to win the prize if they stick, switch, or does it not matter? Clearly identify the probability space of outcomes and their probabilities you use to model this situation. What is the contestant's probability of winning if he uses the best strategy?

## Problems for Section 17.6

### Practice Problems

**Problem 17.26.**

Bruce Lee, on a movie that didn't go public, is practicing by breaking 5 boards with his fists. He is able to break a board with probability 0.8 —he is practicing with his left fist, that's why it's not 1 —and he breaks each board independently.

**(a)** What is the probability that Bruce breaks exactly 2 out of the 5 boards that are placed before him?

**(b)** What is the probability that Bruce breaks at most 3 out of the 5 boards that are placed before him?

**(c)** What is the expected number of boards Bruce will break?

**Problem 17.27.**
Suppose 120 students take a final exam and the mean of their scores is 90. You have no other information about the students and the exam, *e.g.* you should not assume that the highest possible score is 100. You may, however, assume that exam scores are nonnegative.

**(a)** State the best possible upper bound on the number of students who scored at least 180.

**(b)** Now suppose somebody tells you that the lowest score on the exam is 30. Compute the new best possible upper bound on the number of students who scored at least 180.

**Problem 17.28.**
You want to estimate the fraction $p$ of voters in the nation who will vote to re-elect the current president in the upcoming election. You do this by random sampling (with replacement). Specifically, you select $n$ voters independently and randomly, ask them who they are going to vote for, and use the fraction $P$ of those that say they will vote for the current president as an estimate for $p$.

**(a)** Our theorems about sampling and distributions allow us to calculate how confident we can be that the random variable $P$ takes a value near the constant $p$. This calculation uses some facts about voters and the way they are chosen. Which of the following facts are true?

1. Given a particular voter, the probability of that voter preferring the president is $p$.
2. Given a particular voter, the probability of that voter preferring the President is 1 or 0.
3. The probability that some voter is chosen more than once in the sequence goes to zero as $n$ increases.

4. All voters are equally likely to be selected as the third in our sequence of $n$ choices of voters (assuming $n \geq 3$).
5. The probability that the second voter chosen will favor the President, given that the first voter chosen prefers the President, is greater than $p$.
6. The probability that the second voter chosen will favor the President, given that the second voter chosen is from the same state as the first, may not equal $p$.

**(b)** Suppose that, according to your calculations the following is true about your polling:

$$\Pr[|P - p| \leq 0.04] \geq 0.95$$

**(c)** You do the asking, you count how many said they will vote for the President, you divide by $n$, and find that $P = 0.53$. You call the President to give him your results. Which of the following are true?

1. Mr. President, $p = 0.53$!
2. Mr. President, with probability at least 95%, $p$ is within 0.04 of 0.53.
3. Mr. President, either $p$ is within 0.04 of 0.53 or something very strange (5-in-100) has happened.
4. Mr. President, we can be 95% confident that $p$ is within 0.04 of 0.53.

## Exam Problems

**Problem 17.29.**
Sally Smart just graduated from high school. She was accepted to three top colleges.

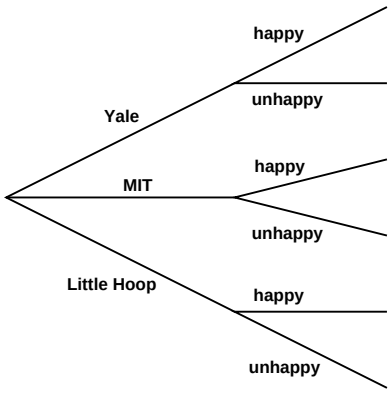
- With probability $4/12$, she attends Yale.
- With probability $5/12$, she attends MIT.
- With probability $3/12$, she attends Little Hoop Community College.

Sally will either be happy or unhappy in college.

- If she attends Yale, she is happy with probability $4/12$.
- If she attends MIT, she is happy with probability $7/12$.

- If she attends Little Hoop, she is happy with probability 11/12.

**(a)** A tree diagram for Sally's situation is shown below. On the diagram, fill in the edge probabilities and at each leaf write the probabilty of that outcome.




**(b)** What is the probability that Sally is happy in college?

**(c)** What is the probability that Sally Smart attends Yale, given that she is happy in college?

**(d)** Show that the event that Sally attends Yale is **not** independent of the event that she is happy.

**(e)** Show that the event that Sally Smart attends MIT **is** independent of the event that she is happy.

**Problem 17.30.**
Construct a probability space $\mathcal{S}$ such that $\mathcal{S}$ contains three events $A$, $B$, and $C$ with the following properties:

- The three events satisfy the "product rule." That is,

$$\Pr[A \cap B \cap C] = \Pr[A] \cdot \Pr[B] \cdot \Pr[C].$$

- The events are *not* mutually independent.

*Hint:* It may be helpful to draw a Venn diagram for $\mathcal{S}$ containing the three events, and then incrementally fill in the probabilities of the disjoint regions.

### Class Problems

**Problem 17.31.**
Let $A, B, C$ be events. For each of the following statements, prove it or give a counterexample.

**(a)** If $A$ is independent of $B$, and $A$ is independent of $C$, then $A$ is independent of $B \cap C$.

**(b)** If $A$ is independent of $B$, and $A$ is independent of $C$, then $A$ is independent of $B \cup C$.

**(c)** If $A$ is independent of $B$, and $A$ is independent of $C$, and $A$ is independent of $B \cap C$, then $A$ is independent of $B \cup C$.

**Problem 17.32.**
Suppose that you flip three fair, mutually independent coins. Define the following events:

- Let $A$ be the event that *the first* coin is heads.
- Let $B$ be the event that *the second* coin is heads.
- Let $C$ be the event that *the third* coin is heads.
- Let $D$ be the event that *an even number of* coins are heads.

**(a)** Use the four step method to determine the probability space for this experiment and the probability of each of $A, B, C, D$.

**(b)** Show that these events are not mutually independent.

**(c)** Show that they are 3-way independent.

### Homework Problems

**Problem 17.33.**
Define the events $A$, $F_{EE}$, $F_{CS}$, $M_{EE}$, and $M_{CS}$ as in Section 17.5.8.

In these terms, the plaintiff in a discrimination suit against a university makes the argument that in both departments, the probability that a woman is granted tenure is less than the probability for a man. That is,

$$\Pr\left[A \mid F_{EE}\right] < \Pr\left[A \mid M_{EE}\right] \quad \text{and} \tag{17.19}$$

$$\Pr\left[A \mid F_{CS}\right] < \Pr\left[A \mid M_{CS}\right]. \tag{17.20}$$

The university's defence attorneys retort that *overall*, a woman applicant is *more* likely to be granted tenure than a man, namely, that

$$\Pr\left[A \mid F_{EE} \cup F_{CS}\right] > \Pr\left[A \mid M_{EE} \cup M_{CS}\right]. \tag{17.21}$$

The judge then interrupts the trial and calls the plaintiff and defence attorneys to a conference in his office to resolve what he thinks are contradictory statements of facts about the tenure data. The judge points out that:

$$\begin{aligned}
&\Pr\left[A \mid F_{EE} \cup F_{CS}\right] \\
&= \Pr\left[A \mid F_{EE}\right] + \Pr\left[A \mid F_{CS}\right] && \text{(because } F_{EE} \text{ and } F_{CS} \text{ are disjoint)} \\
&< \Pr\left[A \mid M_{EE}\right] + \Pr\left[A \mid M_{CS}\right] && \text{(by (17.19) and (17.20))} \\
&= \Pr\left[A \mid M_{EE} \cup M_{CS}\right] && \text{(because } F_{EE} \text{ and } F_{CS} \text{ are disjoint)}
\end{aligned}$$

so

$$\Pr\left[A \mid F_{EE} \cup F_{CS}\right] < \Pr\left[A \mid M_{EE} \cup M_{CS}\right],$$

which directly contradicts the university's position (17.21)!

But the judge is mistaken; an example where the plaintiff and defence assertions are all true appears in Section 17.5.8. What is the mistake in the judge's proof?

**Problem 17.34.**
**Graphs, Logic & Probability**

Let $G$ be an undirected simple graph with $n > 3$ vertices. Let $E(x, y)$ mean that $G$ has an edge between vertices $x$ and $y$, and let $P(x, y)$ mean that there is a length 2 path in $G$ between $x$ and $y$.

**(a)** Explain why $E(x, y)$ implies $P(x, x)$.

**(b)** Circle the mathematical formula that best expresses the definition of $P(x, y)$.

- $P(x, y) ::= \exists z.\ E(x, z)$ AND $E(y, z)$
- $P(x, y) ::= x \neq y$ AND $\exists z.\ E(x, z)$ AND $E(y, z)$
- $P(x, y) ::= \forall z.\ E(x, z)$ OR $E(y, z)$
- $P(x, y) ::= \forall z.\ x \neq y$ IMPLIES $[E(x, z)$ OR $E(y, z)]$

For the following parts (c)–(e), let $V$ be a fixed set of $n > 3$ vertices, and let $G$ be a graph with these vertices constructed randomly as follows: for all distinct vertices

$x, y \in V$, independently include edge $\langle x—y\rangle$ as an edge of $G$ with probability $p$. In particular, $\Pr[E(x, y)] = p$ for all $x \neq y$.

**(c)** For distinct vertices $w$, $x$, $y$ and $z$ in $V$, circle the event pairs that are independent.

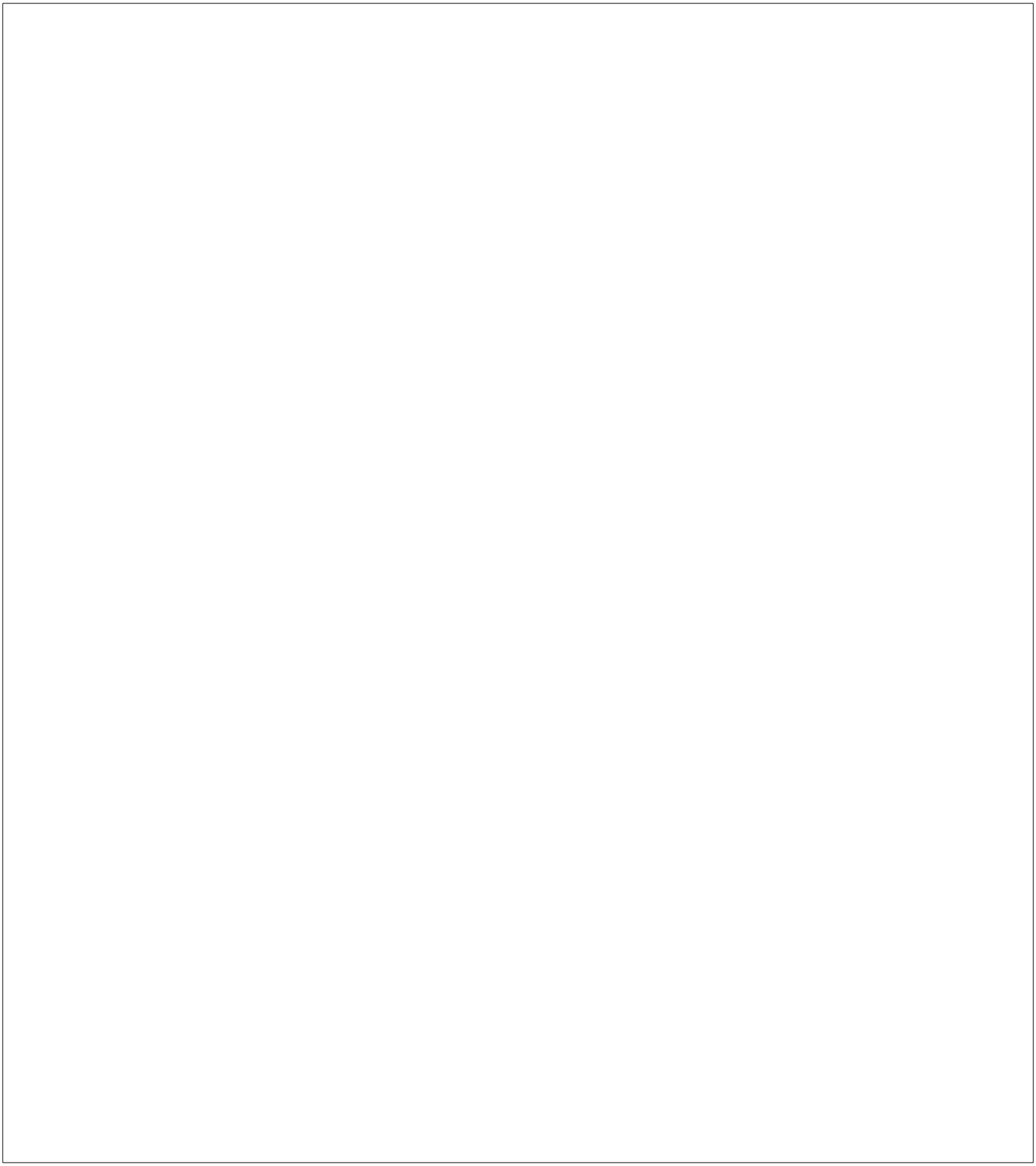
1. $E(w, x)$ versus $E(x, y)$
2. $[E(w, x)$ AND $E(w, y)]$ versus $[E(z, x)$ AND $E(z, y)]$
3. $E(x, y)$ versus $P(x, y)$
4. $P(w, x)$ versus $P(x, y)$
5. $P(w, x)$ versus $P(y, z)$

**(d)** Write a simple formula in terms of $n$ and $p$ for $\Pr[\text{NOT } P(x, y)]$, for distinct vertices $x$ and $y$ in $V$.

*Hint:* Use part (c), item 2.

**(e)** What is the probability that two distinct vertices $x$ and $y$ lie on a three-cycle in $G$? Answer with a simple expression in terms of $p$ and $r$, where $r ::= \Pr[\text{NOT } P(x, y)]$ is the correct answer to part (d).

*Hint:* Express $x$ and $y$ being on a three-cycle as a simple formula involving $E(x, y)$ and $P(x, y)$.



# 18 Random Variables

Thus far, we have focused on probabilities of events. For example, we computed the probability that you win the Monty Hall game or that you have a rare medical condition given that you tested positive. But, in many cases we would like to know more. For example, *how many* contestants must play the Monty Hall game until one of them finally wins? *How long* will this condition last? *How much* will I lose gambling with strange dice all night? To answer such questions, we need to work with random variables.

## 18.1 Random Variable Examples

**Definition 18.1.1.** A *random variable* $R$ on a probability space is a total function whose domain is the sample space.

The codomain of $R$ can be anything, but will usually be a subset of the real numbers. Notice that the name "random variable" is a misnomer; random variables are actually functions!

For example, suppose we toss three independent, unbiased coins. Let $C$ be the number of heads that appear. Let $M = 1$ if the three coins come up all heads or all tails, and let $M = 0$ otherwise. Now every outcome of the three coin flips uniquely determines the values of $C$ and $M$. For example, if we flip heads, tails, heads, then $C = 2$ and $M = 0$. If we flip tails, tails, tails, then $C = 0$ and $M = 1$. In effect, $C$ counts the number of heads, and $M$ indicates whether all the coins match.

Since each outcome uniquely determines $C$ and $M$, we can regard them as functions mapping outcomes to numbers. For this experiment, the sample space is:

$$\mathcal{S} = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}.$$

Now $C$ is a function that maps each outcome in the sample space to a number as follows:

$$\begin{array}{llll} C(HHH) = 3 & & C(THH) = 2 \\ C(HHT) = 2 & & C(THT) = 1 \\ C(HTH) = 2 & & C(TTH) = 1 \\ C(HTT) = 1 & & C(TTT) = 0. \end{array}$$

Similarly, $M$ is a function mapping each outcome another way:

$$\begin{array}{rclrcl} M(HHH) & = & 1 & M(THH) & = & 0 \\ M(HHT) & = & 0 & M(THT) & = & 0 \\ M(HTH) & = & 0 & M(TTH) & = & 0 \\ M(HTT) & = & 0 & M(TTT) & = & 1. \end{array}$$

So $C$ and $M$ are random variables.

### 18.1.1 Indicator Random Variables

An *indicator random variable* is a random variable that maps every outcome to either 0 or 1. Indicator random variables are also called *Bernoulli variables*. The random variable $M$ is an example. If all three coins match, then $M = 1$; otherwise, $M = 0$.

Indicator random variables are closely related to events. In particular, an indicator random variable partitions the sample space into those outcomes mapped to 1 and those outcomes mapped to 0. For example, the indicator $M$ partitions the sample space into two blocks as follows:

$$\underbrace{HHH \quad TTT}_{M=1} \quad \underbrace{HHT \quad HTH \quad HTT \quad THH \quad THT \quad TTH}_{M=0}.$$

In the same way, an event $E$ partitions the sample space into those outcomes in $E$ and those not in $E$. So $E$ is naturally associated with an indicator random variable, $I_E$, where $I_E(\omega) = 1$ for outcomes $\omega \in E$ and $I_E(\omega) = 0$ for outcomes $\omega \notin E$. Thus, $M = I_E$ where $E$ is the event that all three coins match.

### 18.1.2 Random Variables and Events

There is a strong relationship between events and more general random variables as well. A random variable that takes on several values partitions the sample space into several blocks. For example, $C$ partitions the sample space as follows:

$$\underbrace{TTT}_{C=0} \quad \underbrace{TTH \quad THT \quad HTT}_{C=1} \quad \underbrace{THH \quad HTH \quad HHT}_{C=2} \quad \underbrace{HHH}_{C=3}.$$

Each block is a subset of the sample space and is therefore an event. So the assertion that $C = 2$ defines the event

$$[C = 2] = \{THH, HTH, HHT\},$$

and this event has probability

$$\Pr[C = 2] = \Pr[THH] + \Pr[HTH] + \Pr[HHT] = \frac{1}{8} + \frac{1}{8} + \frac{1}{8} = 3/8.$$

Likewise $[M = 1]$ is the event $\{TTT, HHH\}$ and has probability $1/4$.

More generally, any assertion about the values of random variables defines an event. For example, the assertion that $C \leq 1$ defines

$$[C \leq 1] = \{TTT, TTH, THT, HTT\},$$

and so $\Pr[C \leq 1] = 1/2$.

Another example is the assertion that $C \cdot M$ is an odd number. This is an obscure way of saying that all three coins came up heads, namely,

$$[C \cdot M \text{ is odd }] = \{HHH\}.$$

Think about it!

## 18.2 Independence

The notion of independence carries over from events to random variables as well. Random variables $R_1$ and $R_2$ are *independent* iff for all $x_1, x_2$, the two events

$$[R_1 = x_1] \quad \text{and} \quad [R_2 = x_2]$$

are independent.

For example, are $C$ and $M$ independent? Intuitively, the answer should be "no." The number of heads, $C$, completely determines whether all three coins match; that is, whether $M = 1$. But, to verify this intuition, we must find some $x_1, x_2 \in \mathbb{R}$ such that:

$$\Pr[C = x_1 \text{ AND } M = x_2] \neq \Pr[C = x_1] \cdot \Pr[M = x_2].$$

One appropriate choice of values is $x_1 = 2$ and $x_2 = 1$. In this case, we have:

$$\Pr[C = 2 \text{ AND } M = 1] = 0 \neq \frac{1}{4} \cdot \frac{3}{8} = \Pr[M = 1] \cdot \Pr[C = 2].$$

The first probability is zero because we never have exactly two heads ($C = 2$) when all three coins match ($M = 1$). The other two probabilities were computed earlier.

On the other hand, let $H_1$ be the indicator variable for event that the first flip is a Head, so

$$[H_1 = 1] = \{HHH, HTH, HHT, HTT\}.$$

Then $H_1$ is independent of $M$, since

$$\Pr[M = 1] = 1/4 = \Pr\left[M = 1 \mid H_1 = 1\right] = \Pr\left[M = 1 \mid H_1 = 0\right]$$
$$\Pr[M = 0] = 3/4 = \Pr\left[M = 0 \mid H_1 = 1\right] = \Pr\left[M = 0 \mid H_1 = 0\right]$$

This example is an instance of:

**Lemma 18.2.1.** *Two events are independent iff their indicator variables are independent.*

The simple proof is left to Problem 18.1.

As with events, the notion of independence generalizes to more than two random variables.

**Definition 18.2.2.** Random variables $R_1, R_2, \ldots, R_n$ are *mutually independent* iff for all $x_1, x_2, \ldots, x_n$, the $n$ events

$$[R_1 = x_1], [R_2 = x_2], \ldots, [R_n = x_n]$$

are mutually independent.

## 18.3 Distribution Functions

A random variable maps outcomes to values. The probability density function, $\mathrm{PDF}_R(x)$, of a random variable, $R$, measures the probability that $R$ takes the value $x$, and the closely related cumulative distribution function, $\mathrm{CDF}_R(x)$, measures the probability that $R \leq x$. Random variables that show up for different spaces of outcomes often wind up behaving in much the same way because they have the same probability of taking different values, that is, because they have same pdf/cdf.

**Definition 18.3.1.** Let $R$ be a random variable with codomain $V$. The *probability density function* of $R$ is a function $\mathrm{PDF}_R : V \to [0, 1]$ defined by:

$$\mathrm{PDF}_R(x) ::= \begin{cases} \Pr[R = x] & \text{if } x \in \text{range}(R), \\ 0 & \text{if } x \notin \text{range}(R). \end{cases}$$

If the codomain is a subset of the real numbers, then the *cumulative distribution function* is the function $\mathrm{CDF}_R : \mathbb{R} \to [0, 1]$ defined by:

$$\mathrm{CDF}_R(x) ::= \Pr[R \leq x].$$

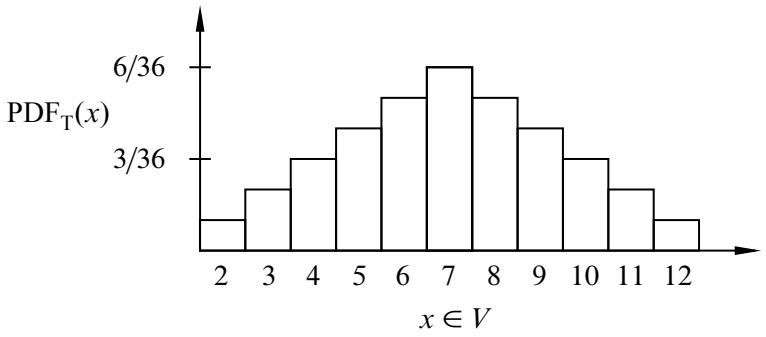



**Figure 18.1** The probability density function for the sum of two 6-sided dice.

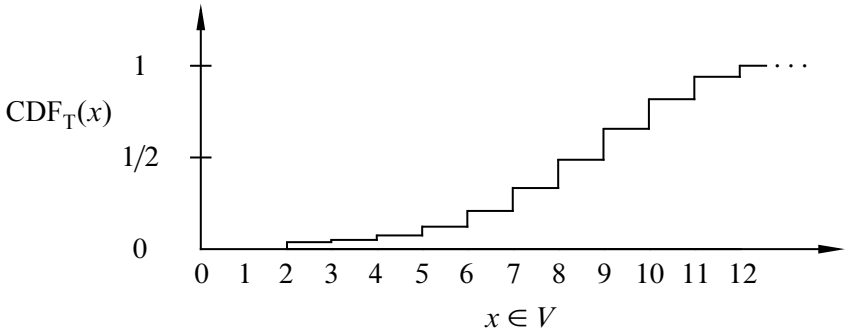



**Figure 18.2** The cumulative distribution function for the sum of two 6-sided dice.

A consequence of this definition is that

$$\sum_{x \in \text{range}(R)} \text{PDF}_R(x) = 1.$$

This is because $R$ has a value for each outcome, so summing the probabilities over all outcomes is the same as summing over the probabilities of each value in the range of $R$.

As an example, suppose that you roll two unbiased, independent, 6-sided dice. Let $T$ be the random variable that equals the sum of the two rolls. This random variable takes on values in the set $V = \{2, 3, \ldots, 12\}$. A plot of the probability density function for $T$ is shown in Figure 18.1. The lump in the middle indicates that sums close to 7 are the most likely. The total area of all the rectangles is 1 since the dice must take on exactly one of the sums in $V = \{2, 3, \ldots, 12\}$.

The cumulative distribution function for $T$ is shown in Figure 18.2: The height of the $i$th bar in the cumulative distribution function is equal to the *sum* of the

heights of the leftmost $i$ bars in the probability density function. This follows from the definitions of pdf and cdf:

$$\mathrm{CDF}_R(x) = \Pr[R \leq x] = \sum_{y \leq x} \Pr[R = y] = \sum_{y \leq x} \mathrm{PDF}_R(y).$$

It also follows from the definition that

$$\lim_{x \to \infty} \mathrm{CDF}_R(x) = 1 \text{ and } \lim_{x \to -\infty} \mathrm{CDF}_R(x) = 0.$$

Both $\mathrm{PDF}_R$ and $\mathrm{CDF}_R$ capture the same information $R$ —obviously each one determines the other —but sometimes one is more convenient. The key point here is that neither the probability density function nor the cumulative distribution function involves the sample space of an experiment.

One of the really interesting things about density functions and distribution functions is that many random variables turn out to have the *same* pdf and cdf. In other words, even though $R$ and $S$ are different random variables on different probability spaces, it is often the case that

$$\mathrm{PDF}_R = \mathrm{PDF}_S.$$

In fact, some pdf's are so common that they are given special names. For example, the three most important distributions in computer science are the *Bernoulli distribution*, the *uniform distribution*, and the *binomial distribution*. We look more closely at these common distributions in the next several sections.

### 18.3.1 Bernoulli Distributions

The Bernoulli distribution is the simplest and most common distribution function. That's because it is the distribution function for an indicator random variable. Specifically, the *Bernoulli distribution* has a probability density function of the form $f_p : \{0, 1\} \to [0, 1]$ where

$$\begin{aligned} f_p(0) &= p, \quad \text{and} \\ f_p(1) &= 1 - p, \end{aligned}$$

for some $p \in [0, 1]$. The corresponding cumulative distribution function is $F_p : \mathbb{R} \to [0, 1]$ where

$$F_p(x) ::= \begin{cases} 0 & \text{if } x < 0 \\ p & \text{if } 0 \leq x < 1 \\ 1 & \text{if } 1 \leq x. \end{cases}$$

### 18.3.2 Uniform Distributions

A random variable that takes on each possible value in its codomain with the same probability is said to be *uniform*. If the codomain $V$ has $n$ elements, then the *uniform distribution* has a pdf of the form

$$f : V \to [0, 1]$$

where

$$f(v) = \frac{1}{n}$$

for all $v \in V$.

Uniform distributions come up all the time. For example, the number rolled on a fair die is uniform on the set $\{1, 2, \ldots, 6\}$. An indicator variable is uniform when its pdf is $f_{1/2}$.

### 18.3.3 The Numbers Game

Enough definitions —let's play a game! We have two envelopes. Each contains an integer in the range $0, 1, \ldots, 100$, and the numbers are distinct. To win the game, you must determine which envelope contains the larger number. To give you a fighting chance, we'll let you peek at the number in one envelope selected at random. Can you devise a strategy that gives you a better than 50% chance of winning?

For example, you could just pick an envelope at random and guess that it contains the larger number. But this strategy wins only 50% of the time. Your challenge is to do better.

So you might try to be more clever. Suppose you peek in one envelope and see the number 12. Since 12 is a small number, you might guess that the number in the other envelope is larger. But perhaps we've been tricky and put small numbers in *both* envelopes. Then your guess might not be so good!

An important point here is that the numbers in the envelopes may *not* be random. We're picking the numbers and we're choosing them in a way that we think will defeat your guessing strategy. We'll only use randomization to choose the numbers if that serves our purpose, which is making you lose!

#### Intuition Behind the Winning Strategy

Amazingly, there is a strategy that wins more than 50% of the time, regardless of what numbers we put in the envelopes!

Suppose that you somehow knew a number $x$ that was in between the numbers in the envelopes. Now you peek in one envelope and see a number. If it is bigger

than $x$, then you know you're peeking at the higher number. If it is smaller than $x$, then you're peeking at the lower number. In other words, if you know a number $x$ between the numbers in the envelopes, then you are certain to win the game.

The only flaw with this brilliant strategy is that you do *not* know such an $x$. Oh well.

But what if you try to *guess* $x$? There is some probability that you guess correctly. In this case, you win 100% of the time. On the other hand, if you guess incorrectly, then you're no worse off than before; your chance of winning is still 50%. Combining these two cases, your overall chance of winning is better than 50%!

Informal arguments about probability, like this one, often sound plausible, but do not hold up under close scrutiny. In contrast, this argument sounds completely implausible —but is actually correct!

### Analysis of the Winning Strategy

For generality, suppose that we can choose numbers from the set $\{0, 1, \ldots, n\}$. Call the lower number $L$ and the higher number $H$.

Your goal is to guess a number $x$ between $L$ and $H$. To avoid confusing equality cases, you select $x$ at random from among the half-integers:

$$\left\{\frac{1}{2},\ 1\frac{1}{2},\ 2\frac{1}{2},\ \ldots,\ n-\frac{1}{2}\right\}$$

But what probability distribution should you use?

The uniform distribution turns out to be your best bet. An informal justification is that if we figured out that you were unlikely to pick some number —say $50\frac{1}{2}$ —then we'd always put 50 and 51 in the envelopes. Then you'd be unlikely to pick an $x$ between $L$ and $H$ and would have less chance of winning.

After you've selected the number $x$, you peek into an envelope and see some number $T$. If $T > x$, then you guess that you're looking at the larger number. If $T < x$, then you guess that the other number is larger.

All that remains is to determine the probability that this strategy succeeds. We can do this with the usual four step method and a tree diagram.

***Step 1: Find the sample space.***

You either choose $x$ too low ($< L$), too high ($> H$), or just right ($L < x < H$). Then you either peek at the lower number ($T = L$) or the higher number ($T = H$). This gives a total of six possible outcomes, as show in Figure 18.3.

***Step 2: Define events of interest.***

The four outcomes in the event that you win are marked in the tree diagram.

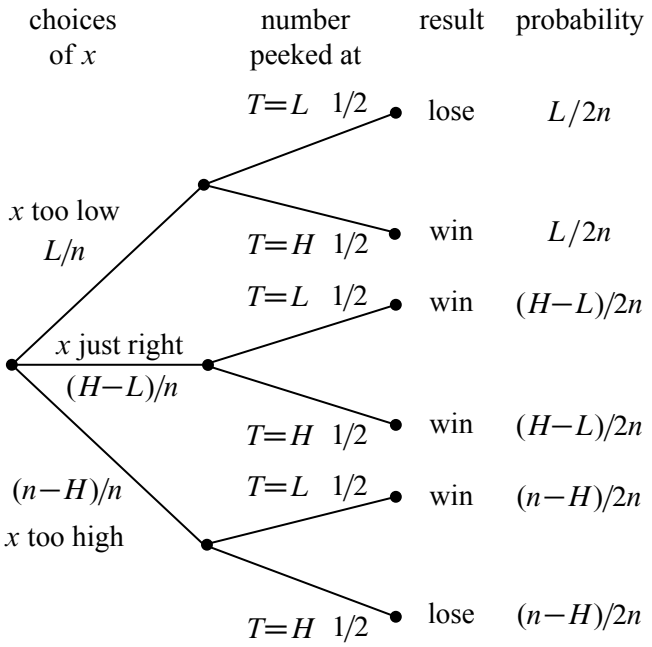



**Figure 18.3** The tree diagram for the numbers game.

***Step 3: Assign outcome probabilities.***
First, we assign edge probabilities. Your guess $x$ is too low with probability $L/n$, too high with probability $(n - H)/n$, and just right with probability $(H - L)/n$. Next, you peek at either the lower or higher number with equal probability. Multiplying along root-to-leaf paths gives the outcome probabilities.

***Step 4: Compute event probabilities.***
The probability of the event that you win is the sum of the probabilities of the four outcomes in that event:

$$\begin{aligned}\Pr[\text{win}] &= \frac{L}{2n} + \frac{H-L}{2n} + \frac{H-L}{2n} + \frac{n-H}{2n} \\ &= \frac{1}{2} + \frac{H-L}{2n} \\ &\geq \frac{1}{2} + \frac{1}{2n}\end{aligned}$$

The final inequality relies on the fact that the higher number $H$ is at least 1 greater than the lower number $L$ since they are required to be distinct.

Sure enough, you win with this strategy more than half the time, regardless of the numbers in the envelopes! For example, if I choose numbers in the range

$0, 1, \ldots, 100$, then you win with probability at least $1/2 + 1/200 = 50.5\%$. Even better, if I'm allowed only numbers in the range $0, \ldots, 10$, then your probability of winning rises to 55%! By Las Vegas standards, those are great odds!

### Randomized Algorithms

The best strategy to win the numbers game is an example of a *randomized algorithm*—it uses random numbers to influence decisions. Protocols and algorithms that make use of random numbers are very important in computer science. There are many problems for which the best known solutions are based on a random number generator.

For example, the most commonly-used protocol for deciding when to send a broadcast on a shared bus or Ethernet is a randomized algorithm known as *exponential backoff*. One of the most commonly-used sorting algorithms used in practice, called *quicksort*, uses random numbers. You'll see many more examples if you take an algorithms course. In each case, randomness is used to improve the probability that the algorithm runs quickly or otherwise performs well.

## 18.3.4 Binomial Distributions

The third commonly-used distribution in computer science is the *binomial distribution*. The standard example of a random variable with a binomial distribution is the number of heads that come up in $n$ independent flips of a coin. If the coin is fair, then the number of heads has an *unbiased binomial distribution*, specified by the pdf $f_n : \{0, 1, \ldots, n\} \to [0, 1]$:

$$f_n(k) ::= \binom{n}{k} 2^{-n}.$$

This is because there are $\binom{n}{k}$ sequences of $n$ coin tosses with exactly $k$ heads, and each such sequence has probability $2^{-n}$.

A plot of $f_{20}(k)$ is shown in Figure 18.4. The most likely outcome is $k = 10$ heads, and the probability falls off rapidly for larger and smaller values of $k$. The falloff regions to the left and right of the main hump are called the *tails of the distribution*.

In many fields, including Computer Science, probability analyses come down to getting small bounds on the tails of the binomial distribution. In the context of a problem, this typically means that there is very small probability that something *bad* happens, which could be a server or communication link overloading or a randomized algorithm running for an exceptionally long time or producing the wrong result.

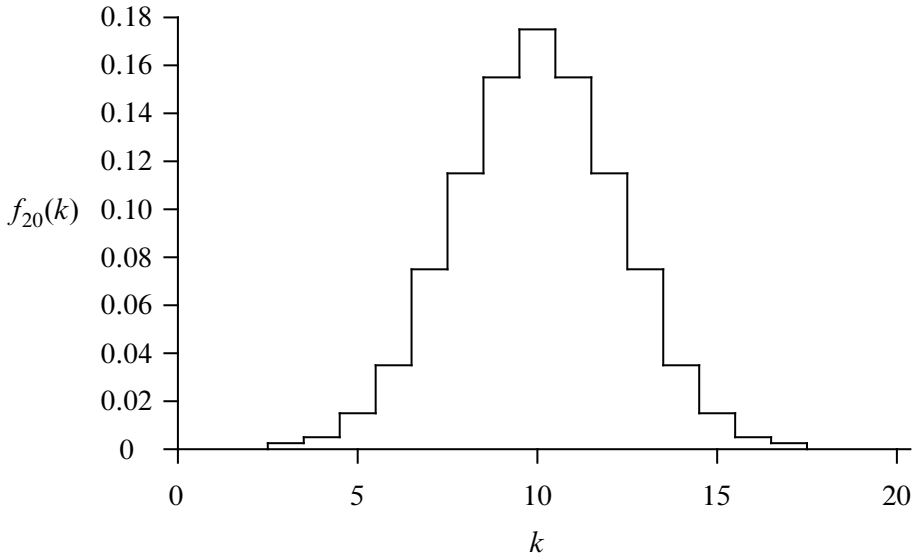



**Figure 18.4** The pdf for the unbiased binomial distribution for $n = 20$, $f_{20}(k)$.

As an example, we can calculate the probability of flipping at most 25 heads in 100 tosses of a fair coin and see that it is very small, namely, less than 1 in 3,000,000.

In fact, the tail of the distribution falls off so rapidly that the probability of flipping exactly 25 heads is nearly twice the probability of flipping fewer than 25 heads! That is, the probability of flipping exactly 25 heads —small as it is — is still nearly twice as large as the probability of flipping exactly 24 heads *plus* the probability of flipping exactly 23 heads *plus* ... the probability of flipping no heads.

### The General Binomial Distribution

If the coins are biased so that each coin is heads with probability $p$, then the number of heads has a *general binomial density function* specified by the pdf

$$f_{n,p} : \{0, 1, \ldots, n\} \to [0, 1]$$

where

$$f_{n,p}(k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

for some $n \in \mathbb{N}^+$ and $p \in [0, 1]$. This is because there are $\binom{n}{k}$ sequences with $k$ heads and $n - k$ tails, but now the probability of each such sequence is $p^k(1-p)^{n-k}$.

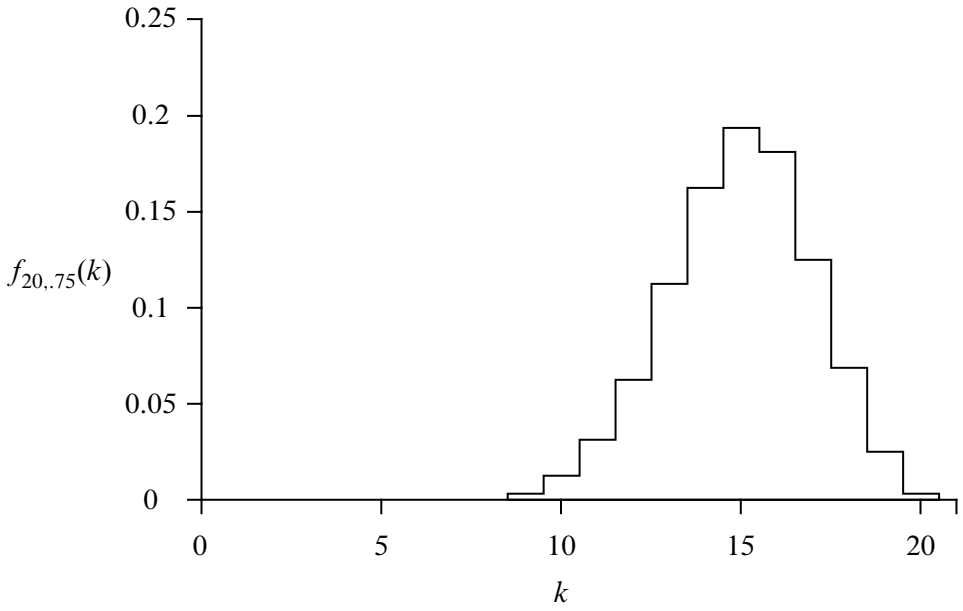



**Figure 18.5** The pdf for the general binomial distribution $f_{n,p}(k)$ for $n = 20$ and $p = .75$.

For example, the plot in Figure 18.5 shows the probability density function $f_{n,p}(k)$ corresponding to flipping $n = 20$ independent coins that are heads with probability $p = 0.75$. The graph shows that we are most likely to get $k = 15$ heads, as you might expect. Once again, the probability falls off quickly for larger and smaller values of $k$.

## 18.4 Great Expectations

The *expectation* or *expected value* of a random variable is a single number that reveals a lot about the behavior of the variable. The expectation of a random variable is also known as its *mean* or *average*. It is the average value of the variable where each value is weighted according to its probability.

For example, suppose we select a student uniformly at random from the class, and let $R$ be the student's quiz score. Then $\mathrm{Ex}[R]$ is just the class average —the first thing everyone wants to know after getting their test back! For similar reasons, the first thing you usually want to know about a random variable is its expected value.

Formally, the expected value of a random variable is defined as follows:

**Definition 18.4.1.** If $R$ is a random variable defined on a sample space $S$, then the expectation of $R$ is

$$\mathrm{Ex}[R] ::= \sum_{\omega \in S} R(\omega) \Pr[\omega]. \tag{18.1}$$

Let's work through some examples.

### 18.4.1 The Expected Value of a Uniform Random Variable

Rolling a 6-sided die provides an example of a uniform random variable. Let $R$ be the value that comes up when you roll a fair 6-sided die. Then by (18.1), the expected value of $R$ is

$$\mathrm{Ex}[R] = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = \frac{7}{2}.$$

This calculation shows that the name "expected" value is a little misleading; the random variable might *never* actually take on that value. You don't ever expect to roll a $3\frac{1}{2}$ on an ordinary die!

In general, if $R_n$ is a random variable with a uniform distribution on $\{1, 2, \ldots, n\}$, then

$$\mathrm{Ex}[R_n] = \sum_{i=1}^{n} i \cdot \frac{1}{n} = \frac{n(n+1)}{2n} = \frac{n+1}{2}.$$

### 18.4.2 The Expected Value of a Reciprocal Random Variable

Define a random variable $S$ to be the reciprocal of the value that comes up when you roll a fair 6-sided die. That is, $S = 1/R$ where $R$ is the value that you roll. Now,

$$\mathrm{Ex}[S] = \mathrm{Ex}\left[\frac{1}{R}\right] = \frac{1}{1} \cdot \frac{1}{6} + \frac{1}{2} \cdot \frac{1}{6} + \frac{1}{3} \cdot \frac{1}{6} + \frac{1}{4} \cdot \frac{1}{6} + \frac{1}{5} \cdot \frac{1}{6} + \frac{1}{6} \cdot \frac{1}{6} = \frac{49}{120}.$$

Notice that

$$\mathrm{Ex}\left[1/R\right] \neq 1/\mathrm{Ex}[R].$$

Assuming that these two quantities are equal is a common mistake.

### 18.4.3 The Expected Value of an Indicator Random Variable

The expected value of an indicator random variable for an event is just the probability of that event.

**Lemma 18.4.2.** *If $I_A$ is the indicator random variable for event $A$, then*

$$\mathrm{Ex}[I_A] = \Pr[A].$$

*Proof.*

$$\begin{aligned}\mathrm{Ex}[I_A] &= 1 \cdot \Pr[I_A = 1] + 0 \cdot \Pr[I_A = 0] = \Pr[I_A = 1] \\ &= \Pr[A]. \qquad \text{(def of } I_A\text{)}\end{aligned}$$

For example, if $A$ is the event that a coin with bias $p$ comes up heads, then $\mathrm{Ex}[I_A] = \Pr[I_A = 1] = p$.

### 18.4.4 Alternate Definition of Expectation

There is another standard way to define expectation.

**Theorem 18.4.3.** *For any random variable $R$,*

$$\mathrm{Ex}[R] = \sum_{x \in \mathrm{range}(R)} x \cdot \Pr[R = x]. \tag{18.2}$$

The proof of Theorem 18.4.3, like many of the elementary proofs about expectation in this chapter, follows by judicious regrouping of terms in equation (18.1):

*Proof.* Suppose $R$ is defined on a sample space $\mathcal{S}$. Then,

$$\begin{aligned}\mathrm{Ex}[R] ::&= \sum_{\omega \in \mathcal{S}} R(\omega) \Pr[\omega] \\ &= \sum_{x \in \mathrm{range}(R)} \sum_{\omega \in [R=x]} R(\omega) \Pr[\omega] \\ &= \sum_{x \in \mathrm{range}(R)} \sum_{\omega \in [R=x]} x \Pr[\omega] \qquad \text{(def of the event } [R = x]\text{)} \\ &= \sum_{x \in \mathrm{range}(R)} x \left( \sum_{\omega \in [R=x]} \Pr[\omega] \right) \qquad \text{(factoring } x \text{ from the inner sum)} \\ &= \sum_{x \in \mathrm{range}(R)} x \cdot \Pr[R = x]. \qquad \text{(def of } \Pr[R = x]\text{)}\end{aligned}$$

The first equality follows because the events $[R = x]$ for $x \in \mathrm{range}(R)$ partition the sample space $\mathcal{S}$, so summing over the outcomes in $[R = x]$ for $x \in \mathrm{range}(R)$ is the same as summing over $\mathcal{S}$. ■

In general, equation (18.2) is more useful than the defining equation (18.1) for calculating expected values. It also has the advantage that it does not depend on the sample space, but only on the density function of the random variable. On the other hand, summing over all outcomes as in equation (18.1) sometimes yields easier proofs about general properties of expectation.

### Medians

The mean of a random variable is not the same as the *median*. The median is the *midpoint* of a distribution.

**Definition 18.4.4.** The *median* of a random variable $R$ is the value $x \in \text{range}(R)$ such that

$$\Pr[R \leq x] \leq \frac{1}{2} \qquad \text{and}$$
$$\Pr[R > x] < \frac{1}{2}.$$

We won't devote much attention to the median. The expected value is more useful and has much more interesting properties.

## 18.4.5 Conditional Expectation

Just like event probabilities, expectations can be conditioned on some event. Given a random variable $R$, the expected value of $R$ conditioned on an event $A$ is the probability-weighted average value of $R$ over outcomes in $A$. More formally:

**Definition 18.4.5.** The *conditional expectation* $\text{Ex}[R \mid A]$ of a random variable $R$ given event $A$ is:

$$\text{Ex}[R \mid A] ::= \sum_{r \in \text{range}(R)} r \cdot \Pr\left[R = r \mid A\right]. \tag{18.3}$$

For example, we can compute the expected value of a roll of a fair die, given that the number rolled is at least 4. We do this by letting $R$ be the outcome of a roll of the die. Then by equation (18.3),

$$\text{Ex}[R \mid R \geq 4] = \sum_{i=1}^{6} i \cdot \Pr\left[R = i \mid R \geq 4\right] = 1{\cdot}0 + 2{\cdot}0 + 3{\cdot}0 + 4{\cdot}\tfrac{1}{3} + 5{\cdot}\tfrac{1}{3} + 6{\cdot}\tfrac{1}{3} = 5.$$

Conditional expectation is useful in dividing complicated expectation calculations into simpler cases. We can find a desired expectation by calculating the conditional expectation in each simple case and averaging them, weighing each case by its probability.

For example, suppose that 49.8% of the people in the world are male and the rest female —which is more or less true. Also suppose the expected height of a randomly chosen male is $5'\,11''$, while the expected height of a randomly chosen female is $5'\,5.''$ What is the expected height of a randomly chosen person? We can calculate this by averaging the heights of men and women. Namely, let $H$ be the height (in feet) of a randomly chosen person, and let $M$ be the event that the person is male and $F$ the event that the person is female. Then

$$\begin{aligned}\operatorname{Ex}[H] &= \operatorname{Ex}[H \mid M]\Pr[M] + \operatorname{Ex}[H \mid F]\Pr[F] \\ &= (5 + 11/12)\cdot 0.498 + (5 + 5/12)\cdot 0.502 \\ &= 5.665\end{aligned}$$

which is a little less than 5' 8."

This method is justified by:

**Theorem 18.4.6** (Law of Total Expectation)**.** *Let $R$ be a random variable on a sample space $\mathcal{S}$, and suppose that $A_1, A_2, \ldots,$ is a partition of $\mathcal{S}$. Then*

$$\operatorname{Ex}[R] = \sum_i \operatorname{Ex}[R \mid A_i]\Pr[A_i].$$

*Proof.*

$$\begin{aligned}\operatorname{Ex}[R] &= \sum_{r\in \operatorname{range}(R)} r\cdot \Pr[R = r] && \text{(by 18.2)}\\ &= \sum_r r\cdot \sum_i \Pr\left[R = r \mid A_i\right]\Pr[A_i] && \text{(Law of Total Probability)}\\ &= \sum_r \sum_i r\cdot \Pr\left[R = r \mid A_i\right]\Pr[A_i] && \text{(distribute constant } r\text{)}\\ &= \sum_i \sum_r r\cdot \Pr\left[R = r \mid A_i\right]\Pr[A_i] && \text{(exchange order of summation)}\\ &= \sum_i \Pr[A_i] \sum_r r\cdot \Pr\left[R = r \mid A_i\right] && \text{(factor constant } \Pr[A_i]\text{)}\\ &= \sum_i \Pr[A_i]\operatorname{Ex}[R \mid A_i]. && \text{(Def 18.4.5 of cond. expectation)}\end{aligned}$$

■

### 18.4.6 Mean Time to Failure

A computer program crashes at the end of each hour of use with probability $p$, if it has not crashed already. What is the expected time until the program crashes?

This will be easy to figure out using the Law of Total Expectation, Theorem 18.4.6. Specifically, we want to find $\mathrm{Ex}[C]$ where $C$ is the number of hours until the first crash. We'll do this by conditioning on whether or not the crash occurs in the first hour.

So let $A$ to be the event that the system fails on the first step and $\overline{A}$ to be the complementary event that the system does not fail on the first step. Then the mean time to failure $\mathrm{Ex}[C]$ is

$$\mathrm{Ex}[C] = \mathrm{Ex}[C \mid A]\Pr[A] + \mathrm{Ex}[C \mid \overline{A}]\Pr[\overline{A}]. \tag{18.4}$$

Since $A$ is the condition that the system crashes on the first step, we know that

$$\mathrm{Ex}[C \mid A] = 1. \tag{18.5}$$

Since $\overline{A}$ is the condition that the system does *not* crash on the first step, conditioning on $\overline{A}$ is equivalent to taking a first step without failure and then starting over without conditioning. Hence,

$$\mathrm{Ex}[C \mid \overline{A}] = 1 + \mathrm{Ex}[C]. \tag{18.6}$$

Plugging (18.5) and (18.6) into (18.4):

$$\begin{aligned}\mathrm{Ex}[C] &= 1 \cdot p + (1 + \mathrm{Ex}[C])(1 - p)\\ &= p + 1 - p + (1 - p)\,\mathrm{Ex}[C]\\ &= 1 + (1 - p)\,\mathrm{Ex}[C].\end{aligned}$$

Then, rearranging terms gives

$$1 = \mathrm{Ex}[C] - (1 - p)\,\mathrm{Ex}[C] = p\,\mathrm{Ex}[C],$$

and thus

$$\mathrm{Ex}[C] = 1/p.$$

The general principle here is well-worth remembering.

## Mean Time to Failure

If a system independently fails at each time step with probability $p$, then the expected number of steps up to the first failure is $1/p$.

So, for example, if there is a 1% chance that the program crashes at the end of each hour, then the expected time until the program crashes is $1/0.01 = 100$ hours.

As a further example, suppose a couple wants to have a baby girl. For simplicity assume there is a 50% chance that each child they have is a girl, and the genders of their children are mutually independent. If the couple insists on having children until they get a girl, then how many baby boys should they expect first?

This is really a variant of the previous problem. The question, "How many hours until the program crashes?" is mathematically the same as the question, "How many children must the couple have until they get a girl?" In this case, a crash corresponds to having a girl, so we should set $p = 1/2$. By the preceding analysis, the couple should expect a baby girl after having $1/p = 2$ children. Since the last of these will be the girl, they should expect just one boy.

Something to think about: If every couple follows the strategy of having children until they get a girl, what will eventually happen to the fraction of girls born in this world?

Using the Law of Total Expectation to find expectations is a worthwhile approach to keep in mind, but it's good review to derive the same formula directly from the definition of expectation. Namely, the probability that the first crash occurs in the $i$th hour for some $i > 0$ is the probability, $(1-p)^{i-1}$, that it does not crash in each of the first $i-1$ hours, times the probability, $p$, that it does crash in the $i$th hour. So

$$\begin{aligned}
\mathrm{Ex}[C] &= \sum_{i\in\mathbb{N}} i \cdot \Pr[C = i] && \text{(by (18.2))}\\
&= \sum_{i\in\mathbb{N}} i(1-p)^{i-1}p\\
&= \frac{p}{1-p}\cdot\sum_{i\in\mathbb{N}} i(1-p)^{i}. && (18.7)
\end{aligned}$$

But we've already seen a sum like this last one (you did remember this, right?), namely, equation (14.13):

$$\sum_{i\in\mathbb{N}} ix^i = \frac{x}{(1-x)^2}.$$

Combining (14.13) with (18.7) gives

$$\mathrm{Ex}[C] = \frac{p}{1-p}\cdot\frac{1-p}{(1-(1-p))^2} = \frac{1}{p}$$

as expected.

For the record, we'll state a formal version of this result. A random variable like $C$ that counts steps to first failure is said to have a *geometric distribution* with paramter $p$.

**Definition 18.4.7.** A random variable, $C$, has a *geometric distribution* with paramter $p$ iff codomain$(C) = \mathbb{Z}^+$ and

$$\Pr[C = i] = (1-p)^{i-1}\,p.$$

**Lemma 18.4.8.** *If a random variable $C$ had a geometric distribution with paramter $p$, then*

$$\mathrm{Ex}[C] = \frac{1}{p}. \tag{18.8}$$

### 18.4.7 Expected Returns in Gambling Games

Some of the most interesting examples of expectation can be explained in terms of gambling games. For straightforward games where you win $w$ dollars with probability $p$ and you lose $x$ dollars with probability $1-p$, it is easy to compute your *expected return* or *winnings*. It is simply

$$pw - (1-p)x \text{ dollars.}$$

For example, if you are flipping a fair coin and you win \$1 for heads and you lose \$1 for tails, then your expected winnings are

$$\frac{1}{2}\cdot 1 - \left(1-\frac{1}{2}\right)\cdot 1 = 0.$$

In such cases, the game is said to be *fair* since your expected return is zero.

Some gambling games are more complicated and thus more interesting. The following game where the winners split a pot is representative of many poker games, betting pools, and lotteries.

#### Splitting the Pot

After your last encounter with biker dude, one thing led to another and you have dropped out of school and become a Hell's Angel. It's late on a Friday night and, feeling nostalgic for the old days, you drop by your old hangout, where you encounter two of your former TAs, Eric and Nick. Eric and Nick propose that you join them in a simple wager. Each player will put \$2 on the bar and secretly write "heads" or "tails" on their napkin. Then one player will flip a fair coin. The \$6 on the bar will then be divided equally among the players who correctly predicted the outcome of the coin toss.

After your life-altering encounter with strange dice, you are more than a little skeptical. So Eric and Nick agree to let you be the one to flip the coin. This certainly seems fair. How can you lose?

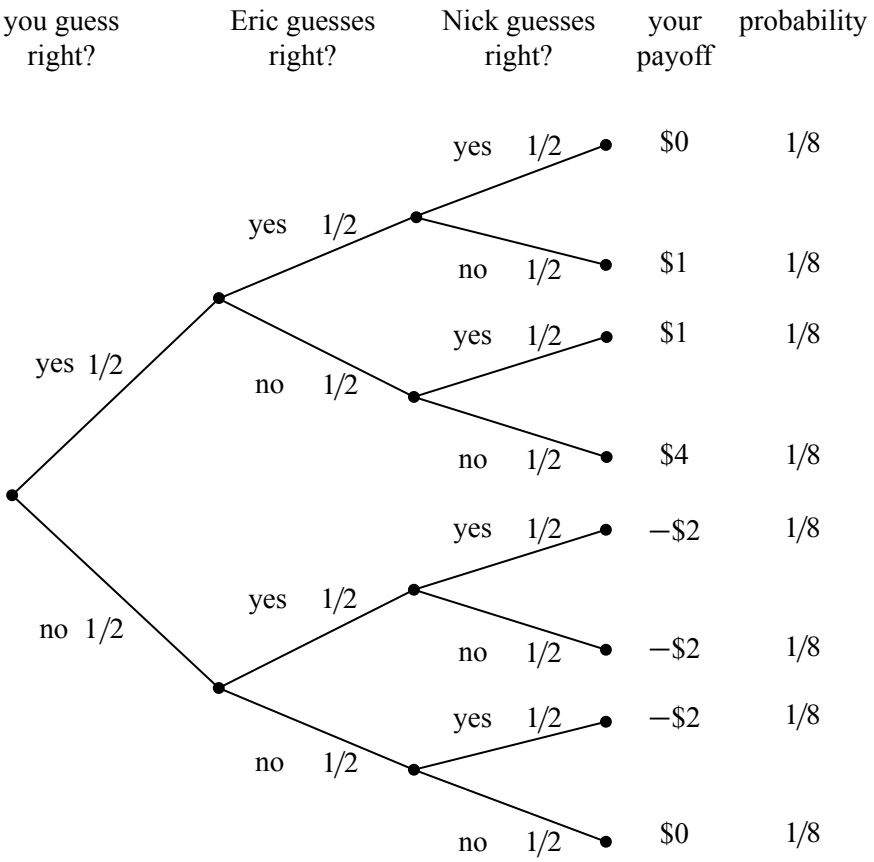



**Figure 18.6** The tree diagram for the game where three players each wager $2 and then guess the outcome of a fair coin toss. The winners split the pot.

But you have learned your lesson and so before agreeing, you go through the four-step method and write out the tree diagram to compute your expected return. The tree diagram is shown in Figure 18.6.

The "payoff" values in Figure 18.6 are computed by dividing the $6 pot[1] among those players who guessed correctly and then subtracting the $2 that you put into the pot at the beginning. For example, if all three players guessed correctly, then your payoff is $0, since you just get back your $2 wager. If you and Nick guess correctly and Eric guessed wrong, then your payoff is

$$\frac{6}{2} - 2 = 1.$$

[1]The money invested in a wager is commonly referred to as the *pot*.

In the case that everyone is wrong, you all agree to split the pot and so, again, your payoff is zero.

To compute your expected return, you use equation (18.2):

$$\begin{aligned}\text{Ex[payoff]} &= 0\cdot\frac{1}{8} + 1\cdot\frac{1}{8} + 1\cdot\frac{1}{8} + 4\cdot\frac{1}{8} \\ &\quad + (-2)\cdot\frac{1}{8} + (-2)\cdot\frac{1}{8} + (-2)\cdot\frac{1}{8} + 0\cdot\frac{1}{8} \\ &= 0.\end{aligned}$$

This confirms that the game is fair. So, for old time's sake, you break your solemn vow to never ever engage in strange gambling games.

### The Impact of Collusion

Needless to say, things are not turning out well for you. The more times you play the game, the more money you seem to be losing. After 1000 wagers, you have lost over $500. As Nick and Eric are consoling you on your "bad luck," you remember how rapidly the tails of the binomial distribute decrease, suggesting that the probability of losing $500 in 1000 fair $2 wagers is less than the probability of being struck by lightning while playing poker and being dealt four Aces. How can this be?

It is possible that you are truly very very unlucky. But it is more likely that something is wrong with the tree diagram in Figure 18.6 and that "something" just might have something to do with the possibility that Nick and Eric are colluding against you.

To be sure, Nick and Eric can only guess the outcome of the coin toss with probability $1/2$, but what if Nick and Eric always guess differently? In other words, what if Nick always guesses "tails" when Eric guesses "heads," and vice-versa? This would result in a slightly different tree diagram, as shown in Figure 18.7.

The payoffs for each outcome are the same in Figures 18.6 and 18.7, but the probabilities of the outcomes are different. For example, it is no longer possible for all three players to guess correctly, since Nick and Eric are always guessing differently. More importantly, the outcome where your payoff is $4 is also no longer possible. Since Nick and Eric are always guessing differently, one of them will always get a share of the pot. As you might imagine, this is not good for you!

When we use equation (18.2) to compute your expected return in the collusion

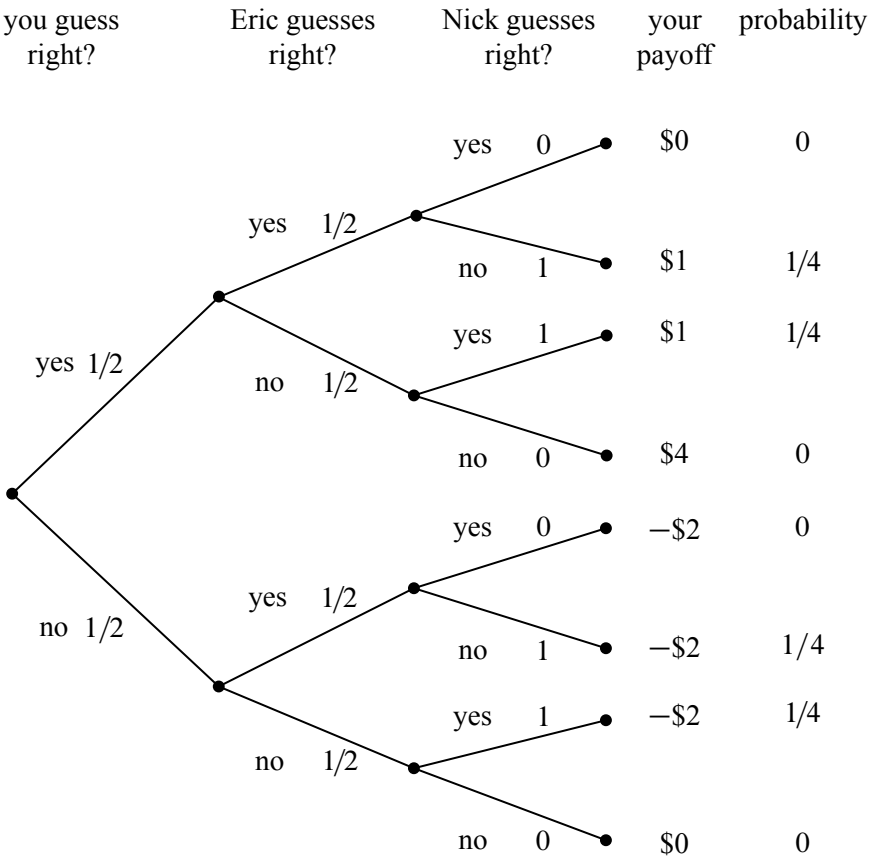



**Figure 18.7** The revised tree diagram reflecting the scenario where Nick always guesses the opposite of Eric.

scenario, we find that

$$\begin{aligned}\text{Ex[payoff]} &= 0\cdot 0 + 1\cdot\frac{1}{4} + 1\cdot\frac{1}{4} + 4\cdot 0 \\ &\quad + (-2)\cdot 0 + (-2)\cdot\frac{1}{4} + (-2)\cdot\frac{1}{4} + 0\cdot 0 \\ &= -\frac{1}{2}.\end{aligned}$$

This is very bad indeed. By colluding, Nick and Eric have made it so that you expect to lose $.50 every time you play. No wonder you lost $500 over the course of 1000 wagers.

Maybe it would be a good idea to go back to school —your Hell's Angels buds may not be too happy that you just lost their $500.

### How to Win the Lottery

Similar opportunities to "collude" arise in many betting games. For example, consider the typical weekly football betting pool, where each participant wagers $10 and the participants that pick the most games correctly split a large pot. The pool seems fair if you think of it as in Figure 18.6. But, in fact, if two or more players collude by guessing differently, they can get an "unfair" advantage at your expense!

In some cases, the collusion is inadvertent and you can profit from it. For example, many years ago, a former MIT Professor of Mathematics named Herman Chernoff figured out a way to make money by playing the state lottery. This was surprising since state lotteries typically have very poor expected returns. That's because the state usually takes a large share of the wagers before distributing the rest of the pot among the winners. Hence, anyone who buys a lottery ticket is expected to *lose* money. So how did Chernoff find a way to make money? It turned out to be easy!

In a typical state lottery,

- all players pay $1 to play and select 4 numbers from 1 to 36,
- the state draws 4 numbers from 1 to 36 uniformly at random,
- the states divides 1/2 of the money collected among the people who guessed correctly and spends the other half redecorating the governor's residence.

This is a lot like the game you played with Nick and Eric, except that there are more players and more choices. Chernoff discovered that a small set of numbers was selected by a large fraction of the population. Apparently many people think the same way; they pick the same numbers not on purpose as in the previous game with Nick and Eric, but based on Manny's batting average or today's date.

It was as if the players were colluding to lose! If any one of them guessed correctly, then they'd have to split the pot with many other players. By selecting numbers uniformly at random, Chernoff was unlikely to get one of these favored sequences. So if he won, he'd likely get the whole pot! By analyzing actual state lottery data, he determined that he could win an average of 7 cents on the dollar. In other words, his expected return was not −$.50 as you might think, but +$.07.[2]

Inadvertent collusion often arises in betting pools and is a phenomenon that you can take advantage of. For example, suppose you enter a Super Bowl betting pool where the goal is to get closest to the total number of points scored in the game. Also suppose that the average Super Bowl has a total of 30 point scored and that everyone knows this. Then most people will guess around 30 points. Where should you guess? Well, you should guess just outside of this range because you get to cover a lot more ground and you don't share the pot if you win. Of course, if you are in a pool with math students and they all know this strategy, then maybe you should guess 30 points after all.

## 18.5 Linearity of Expectation

Expected values obey a simple, very helpful rule called *Linearity of Expectation*. Its simplest form says that the expected value of a sum of random variables is the sum of the expected values of the variables.

**Theorem 18.5.1.** *For any random variables $R_1$ and $R_2$,*

$$\mathrm{Ex}[R_1 + R_2] = \mathrm{Ex}[R_1] + \mathrm{Ex}[R_2].$$

*Proof.* Let $T ::= R_1 + R_2$. The proof follows straightforwardly by rearranging terms in equation (18.1) in the definition of expectation:

$$\begin{aligned}
\mathrm{Ex}[T] &::= \sum_{\omega \in S} T(\omega) \cdot \Pr[\omega] \\
&= \sum_{\omega \in S} (R_1(\omega) + R_2(\omega)) \cdot \Pr[\omega] && \text{(def of } T\text{)} \\
&= \sum_{\omega \in S} R_1(\omega) \Pr[\omega] + \sum_{\omega \in S} R_2(\omega) \Pr[\omega] && \text{(rearranging terms)} \\
&= \mathrm{Ex}[R_1] + \mathrm{Ex}[R_2]. && \text{(by (18.1))}
\end{aligned}$$

■

[2] Most lotteries now offer randomized tickets to help smooth out the distribution of selected sequences.

A small extension of this proof, which we leave to the reader, implies

**Theorem 18.5.2.** *For random variables $R_1$, $R_2$ and constants $a_1, a_2 \in \mathbb{R}$,*

$$\mathrm{Ex}[a_1 R_1 + a_2 R_2] = a_1 \mathrm{Ex}[R_1] + a_2 \mathrm{Ex}[R_2].$$

In other words, expectation is a linear function. A routine induction extends the result to more than two variables:

**Corollary 18.5.3** (Linearity of Expectation)**.** *For any random variables $R_1, \ldots, R_k$ and constants $a_1, \ldots, a_k \in \mathbb{R}$,*

$$\mathrm{Ex}\left[\sum_{i=1}^{k} a_i R_i\right] = \sum_{i=1}^{k} a_i \mathrm{Ex}[R_i].$$

The great thing about linearity of expectation is that *no independence is required*. This is really useful, because dealing with independence is a pain, and we often need to work with random variables that are not known to be independent.

As an example, let's compute the expected value of the sum of two fair dice.

### 18.5.1 Expected Value of Two Dice

What is the expected value of the sum of two fair dice?

Let the random variable $R_1$ be the number on the first die, and let $R_2$ be the number on the second die. We observed earlier that the expected value of one die is 3.5. We can find the expected value of the sum using linearity of expectation:

$$\mathrm{Ex}[R_1 + R_2] = \mathrm{Ex}[R_1] + \mathrm{Ex}[R_2] = 3.5 + 3.5 = 7.$$

Notice that we did *not* have to assume that the two dice were independent. The expected sum of two dice is 7, even if they are glued together (provided each individual die remains fair after the gluing). Proving that this expected sum is 7 with a tree diagram would be a bother: there are 36 cases. And if we did not assume that the dice were independent, the job would be really tough!

### 18.5.2 Sums of Indicator Random Variables

Linearity of expectation is especially useful when you have a sum of indicator random variables. As an example, suppose there is a dinner party where $n$ men check their hats. The hats are mixed up during dinner, so that afterward each man receives a random hat. In particular, each man gets his own hat with probability $1/n$. What is the expected number of men who get their own hat?

Letting $G$ be the number of men that get their own hat, we want to find the expectation of $G$. But all we know about $G$ is that the probability that a man gets his own hat back is $1/n$. There are many different probability distributions of hat permutations with this property, so we don't know enough about the distribution of $G$ to calculate its expectation directly. But linearity of expectation makes the problem really easy.

The trick[3] is to express $G$ as a sum of indicator variables. In particular, let $G_i$ be an indicator for the event that the $i$th man gets his own hat. That is, $G_i = 1$ if the $i$th man gets his own hat, and $G_i = 0$ otherwise. The number of men that get their own hat is then the sum of these indicator random variables:

$$G = G_1 + G_2 + \cdots + G_n. \tag{18.9}$$

These indicator variables are *not* mutually independent. For example, if $n - 1$ men all get their own hats, then the last man is certain to receive his own hat. But, since we plan to use linearity of expectation, we don't have worry about independence!

Since $G_i$ is an indicator random variable, we know from Lemma 18.4.2 that

$$\mathrm{Ex}[G_i] = \Pr[G_i = 1] = 1/n. \tag{18.10}$$

By Linearity of Expectation and equation (18.9), this means that

$$\begin{aligned}
\mathrm{Ex}[G] &= \mathrm{Ex}[G_1 + G_2 + \cdots + G_n] \\
&= \mathrm{Ex}[G_1] + \mathrm{Ex}[G_2] + \cdots + \mathrm{Ex}[G_n] \\
&= \overbrace{\frac{1}{n} + \frac{1}{n} + \cdots + \frac{1}{n}}^{n} \\
&= 1.
\end{aligned}$$

So even though we don't know much about how hats are scrambled, we've figured out that on average, just one man gets his own hat back!

More generally, Linearity of Expectation provides a very good method for computing the expected number of events that will happen.

**Theorem 18.5.4.** *Given any collection of events $A_1, A_2, \ldots, A_n$, the expected number of events that will occur is*

$$\sum_{i=1}^{n} \Pr[A_i].$$

[3] We are going to use this trick a lot so it is important to understand it.

For example, $A_i$ could be the event that the $i$th man gets the right hat back. But in general, it could be any subset of the sample space, and we are asking for the expected number of events that will contain a random sample point.

*Proof.* Define $R_i$ to be the indicator random variable for $A_i$, where $R_i(\omega) = 1$ if $w \in A_i$ and $R_i(\omega) = 0$ if $w \notin A_i$. Let $R = R_1 + R_2 + \cdots + R_n$. Then

$$\begin{aligned}
\operatorname{Ex}[R] &= \sum_{i=1}^{n} \operatorname{Ex}[R_i] && \text{(by Linearity of Expectation)} \\
&= \sum_{i=1}^{n} \Pr[R_i = 1] && \text{(by Lemma 18.4.2)} \\
&= \sum_{i=1}^{n} \Pr[A_i]. && \text{(def of indicator variable)}
\end{aligned}$$

So whenever you are asked for the expected number of events that occur, all you have to do is sum the probabilities that each event occurs. Independence is not needed.

### 18.5.3 Expectation of a Binomial Distribution

Suppose that we independently flip $n$ biased coins, each with probability $p$ of coming up heads. What is the expected number of heads?

Let $J$ be the random variable denoting the number of heads. Then $J$ has a binomial distribution with parameters $n$, $p$, and

$$\Pr[J = k] = \binom{n}{k} p^k (1-p)^{n-k}.$$

Applying equation (18.2), this means that

$$\operatorname{Ex}[J] = \sum_{k=0}^{n} k \Pr[J = k] = \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k}. \tag{18.11}$$

This sum looks a tad nasty, but linearity of expectation leads to an easy derivation of a simple closed form. We just express $J$ as a sum of indicator random variables, which is easy. Namely, let $J_i$ be the indicator random variable for the $i$th coin coming up heads, that is,

$$J_i ::= \begin{cases} 1 & \text{if the } i\text{th coin is heads} \\ 0 & \text{if the } i\text{th coin is tails.} \end{cases}$$

Then the number of heads is simply

$$J = J_1 + J_2 + \cdots + J_n.$$

By Theorem 18.5.4,

$$\mathrm{Ex}[J] = \sum_{i=1}^{n} \Pr[J_i] = pn. \tag{18.12}$$

That really was easy. If we flip $n$ mutually independent coins, we expect to get $pn$ heads. Hence the expected value of a binomial distribution with parameters $n$ and $p$ is simply $pn$.

But what if the coins are not mutually independent? It doesn't matter —the answer is still $pn$ because Linearity of Expectation and Theorem 18.5.4 do not assume any independence.

If you are not yet convinced that Linearity of Expectation and Theorem 18.5.4 are powerful tools, consider this: without even trying, we have used them to prove a complicated looking identity, namely,

$$\sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k} = pn, \tag{18.13}$$

which follows by combining equations (18.11) and (18.12).[4]

The next section has an even more convincing illustration of the power of linearity to solve a challenging problem.

[4]Equation (18.13) may look daunting initially, but it is, after all, pretty similar to the binomial identity, and that connection leads to a simple derivation by algebra. Namely, starting with the binomial identity

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}.$$

we can differentiate with respect to $x$ (as in Section 14.1.6) to get

$$n(x+y)^{n-1} = \sum_{k=0}^{n} k \binom{n}{k} x^{k-1} y^{n-k}.$$

Multiplying both sides by $x$ gives

$$xn(x+y)^{n-1} = \sum_{k=0}^{n} k \binom{n}{k} x^k y^{n-k} \tag{18.14}$$

Plugging $p$ for $x$ and $1-p$ for $y$ in (18.14) then yields (18.13).

### 18.5.4 The Coupon Collector Problem

Every time we purchase a kid's meal at Taco Bell, we are graciously presented with a miniature "Racin' Rocket" car together with a launching device which enables us to project our new vehicle across any tabletop or smooth floor at high velocity. Truly, our delight knows no bounds.

There are $n$ different types of Racin' Rocket cars (blue, green, red, gray, etc.). The type of car awarded to us each day by the kind woman at the Taco Bell register appears to be selected uniformly and independently at random. What is the expected number of kid's meals that we must purchase in order to acquire at least one of each type of Racin' Rocket car?

The same mathematical question shows up in many guises: for example, what is the expected number of people you must poll in order to find at least one person with each possible birthday? Here, instead of collecting Racin' Rocket cars, you're collecting birthdays. The general question is commonly called the *coupon collector problem* after yet another interpretation.

A clever application of linearity of expectation leads to a simple solution to the coupon collector problem. Suppose there are five different types of Racin' Rocket cars, and we receive this sequence:

$$\text{blue} \quad \text{green} \quad \text{green} \quad \text{red} \quad \text{blue} \quad \text{orange} \quad \text{blue} \quad \text{orange} \quad \text{gray}.$$

Let's partition the sequence into 5 segments:

$$\underbrace{\text{blue}}_{X_0} \quad \underbrace{\text{green}}_{X_1} \quad \underbrace{\text{green} \quad \text{red}}_{X_2} \quad \underbrace{\text{blue} \quad \text{orange}}_{X_3} \quad \underbrace{\text{blue} \quad \text{orange} \quad \text{gray}}_{X_4}.$$

The rule is that a segment ends whenever we get a new kind of car. For example, the middle segment ends when we get a red car for the first time. In this way, we can break the problem of collecting every type of car into stages. Then we can analyze each stage individually and assemble the results using linearity of expectation.

Let's return to the general case where we're collecting $n$ Racin' Rockets. Let $X_k$ be the length of the $k$th segment. The total number of kid's meals we must purchase to get all $n$ Racin' Rockets is the sum of the lengths of all these segments:

$$T = X_0 + X_1 + \cdots + X_{n-1}$$

Now let's focus our attention on $X_k$, the length of the $k$th segment. At the beginning of segment $k$, we have $k$ different types of car, and the segment ends when we acquire a new type. When we own $k$ types, each kid's meal contains a type that we already have with probability $k/n$. Therefore, each meal contains a new type of car with probability $1 - k/n = (n-k)/n$. Thus, the expected number

of meals until we get a new kind of car is $n/(n-k)$ by the Mean Time to Failure rule. This means that

$$\mathrm{Ex}[X_k] = \frac{n}{n-k}.$$

Linearity of expectation, together with this observation, solves the coupon collector problem:

$$\begin{aligned}
\mathrm{Ex}[T] &= \mathrm{Ex}[X_0 + X_1 + \cdots + X_{n-1}] \\
&= \mathrm{Ex}[X_0] + \mathrm{Ex}[X_1] + \cdots + \mathrm{Ex}[X_{n-1}] \\
&= \frac{n}{n-0} + \frac{n}{n-1} + \cdots + \frac{n}{3} + \frac{n}{2} + \frac{n}{1} \\
&= n\left(\frac{1}{n} + \frac{1}{n-1} + \cdots + \frac{1}{3} + \frac{1}{2} + \frac{1}{1}\right) \\
&= n\left(\frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n-1} + \frac{1}{n}\right) \\
&= nH_n \qquad (18.15) \\
&\sim n\ln n.
\end{aligned}$$

Wow! It's those Harmonic Numbers again!

We can use equation (18.15) to answer some concrete questions. For example, the expected number of die rolls required to see every number from 1 to 6 is:

$$6H_6 = 14.7\ldots.$$

And the expected number of people you must poll to find at least one person with each possible birthday is:

$$365H_{365} = 2364.6\ldots.$$

### 18.5.5 Infinite Sums

Linearity of expectation also works for an infinite number of random variables provided that the variables satisfy some stringent absolute convergence criteria.

**Theorem 18.5.5** (Linearity of Expectation)**.** *Let $R_0$, $R_1$, ..., be random variables such that*

$$\sum_{i=0}^{\infty} \mathrm{Ex}[\,|R_i|\,]$$

*converges. Then*

$$\operatorname{Ex}\left[\sum_{i=0}^{\infty} R_i\right] = \sum_{i=0}^{\infty} \operatorname{Ex}[R_i].$$

*Proof.* Let $T ::= \sum_{i=0}^{\infty} R_i$.

We leave it to the reader to verify that, under the given convergence hypothesis, all the sums in the following derivation are absolutely convergent, which justifies rearranging them as follows:

$$\begin{aligned}
\sum_{i=0}^{\infty} \operatorname{Ex}[R_i] &== \sum_{i=0}^{\infty} \sum_{s\in\mathcal{S}} R_i(s)\cdot \Pr[s] && \text{(Def. 18.4.1)}\\
&= \sum_{s\in\mathcal{S}} \sum_{i=0}^{\infty} R_i(s)\cdot \Pr[s] && \text{(exchanging order of summation)}\\
&= \sum_{s\in\mathcal{S}} \left[\sum_{i=0}^{\infty} R_i(s)\right]\cdot \Pr[s] && \text{(factoring out } \Pr[s])\\
&= \sum_{s\in\mathcal{S}} T(s)\cdot \Pr[s] && \text{(Def. of } T)\\
&= \operatorname{Ex}[T] && \text{(Def. 18.4.1)}\\
&= \operatorname{Ex}\left[\sum_{i=0}^{\infty} R_i\right]. && \text{(Def. of } T). \ \blacksquare
\end{aligned}$$

### 18.5.6 Expectations of Products

While the expectation of a sum is the sum of the expectations, the same is usually not true for products. For example, suppose that we roll a fair 6-sided die and denote the outcome with the random variable $R$. Does $\operatorname{Ex}[R\cdot R] = \operatorname{Ex}[R]\cdot\operatorname{Ex}[R]$?

We know that $\operatorname{Ex}[R] = 3\frac{1}{2}$ and thus $\operatorname{Ex}[R]^2 = 12\frac{1}{4}$. Let's compute $\operatorname{Ex}[R^2]$ to see if we get the same result.

$$\begin{aligned}
\operatorname{Ex}\left[R^2\right] &= \sum_{\omega\in\mathcal{S}} R^2(\omega)\Pr[w] = \sum_{i=1}^{6} i^2\cdot\Pr[R_i = i]\\
&= \frac{1^2}{6}+\frac{2^2}{6}+\frac{3^2}{6}+\frac{4^2}{6}+\frac{5^2}{6}+\frac{6^2}{6} = 15\ 1/6 \neq 12\ 1/4.
\end{aligned}$$

That is,

$$\operatorname{Ex}[R\cdot R] \neq \operatorname{Ex}[R]\cdot\operatorname{Ex}[R].$$

So the expectation of a product is not always equal to the product of the expectations.

There is a special case when such a relationship *does* hold however; namely, when the random variables in the product are *independent*.

**Theorem 18.5.6.** *For any two* independent *random variables* $R_1$, $R_2$,

$$\mathrm{Ex}[R_1 \cdot R_2] = \mathrm{Ex}[R_1] \cdot \mathrm{Ex}[R_2].$$

The proof follows by judicious rearrangement of terms in the sum that defines $\mathrm{Ex}[R_1 \cdot R_2]$. Details appear in Problem 18.17.

Theorem 18.5.6 extends routinely to a collection of mutually independent variables.

**Corollary 18.5.7.** *[Expectation of Independent Product]*
*If random variables* $R_1, R_2, \ldots, R_k$ *are mutually independent, then*

$$\mathrm{Ex}\left[\prod_{i=1}^{k} R_i\right] = \prod_{i=1}^{k} \mathrm{Ex}[R_i].$$

## Problems for Section 18.2

### Practice Problems

**Problem 18.1.** **(a)** Prove that if $A$ and $B$ are independent events, then so are $A$ and $\overline{B}$.

**(b)** Let $I_A$ and $I_B$ be the indicator variables for events $A$ and $B$. Prove that $I_A$ and $I_B$ are independent iff $A$ and $B$ are independent.

*Hint:* For any event, $E$, let $E^1 ::= E$ and $E^0 ::= \overline{E}$. So the event $[I_E = a]$ is the same as $E^a$.

### Homework Problems

**Problem 18.2.**
Let $R$, $S$, and $T$ be random variables with the same codomain, $V$.

**(a)** Suppose $R$ is uniform —that is,

$$\Pr[R = b] = \frac{1}{|V|},$$

for all $b \in V$ —and $R$ is independent of $S$. Originally this text had the following argument:

The probability that $R = S$ is the same as the probability that $R$ takes whatever value $S$ happens to have, therefore

$$\Pr[R = S] = \frac{1}{|V|}. \tag{18.16}$$

Are you convinced by this argument? Write out a careful proof of (18.16).

*Hint:* The event $[R = S]$ is a disjoint union of events

$$[R = S] = \bigcup_{b \in V} [R = b \text{ AND } S = b].$$

**(b)** Let $S \times T$ be the random variable giving the values of $S$ and $T$.[5] Now suppose $R$ has a uniform distribution, and $R$ is independent of $S \times T$. How about this argument?

> The probability that $R = S$ is the same as the probability that $R$ equals the first coordinate of whatever value $S \times T$ happens to have, and this probability remains equal to $1/|V|$ by independence. Therefore the event $[R = S]$ is independent of $[S = T]$.

Write out a careful proof that $[R = S]$ is independent of $[S = T]$.

**(c)** Let $V = \{1, 2, 3\}$ and $R, S, T$ take the following values with equal probability,

$$111, 211, 123, 223, 132, 232.$$

Verify that

1. $R$ is independent of $S \times T$,
2. The event $[R = S]$ is not independent of $[S = T]$.
3. $S$ and $T$ have a uniform distribution,

**Problem 18.3.**

Let $R$, $S$, and $T$ be mutually independent random variables with the same codomain, $V$. Problem 18.2 showed that if $R$ is uniform —that is,

$$\Pr[R = b] = \frac{1}{|V|},$$

[5] That is, $S \times T : \mathcal{S} \to V \times V$ where

$$(S \times T)(\omega) ::= (S(\omega), T(\omega))$$

for every outcome $\omega \in \mathcal{S}$.

for all $b \in V$, then

the events $[R = S]$ and $[S = T]$ are independent.

This implies that these events are also independent if $T$ is uniform, since $R$ and $T$ are symmetric in this assertion. Prove converssely that if neither $R$ nor $T$ is uniform, then these events are not independent.

## Problems for Section 18.3

### Practice Problems

**Problem 18.4.**
Suppose $X_1$, $X_2$, and $X_3$ are three mutually independent random variables, each having the uniform distribution

$$\forall k,\ k \in \{1, 2, 3\}.\ \Pr[X_i = k] = \frac{1}{3}.$$

Let $M$ be another random variable giving the maximum of these three random variables. What is the probability density function of $M$?

### Class Problems

Guess the Bigger Number Game

Team 1:

- Write different integers between 0 and 7 on two pieces of paper.
- Put the papers face down on a table.

Team 2:

- Turn over one paper and look at the number on it.
- Either stick with this number or switch to the unseen other number.

Team 2 wins if it chooses the larger number; else, Team 1 wins.

**Problem 18.5.**
The analysis in section 18.3.3 implies that Team 2 has a strategy that wins 4/7 of the time no matter how Team 1 plays. Can Team 2 do better? The answer is "no,"

because Team 1 has a strategy that guarantees that it wins at least 3/7 of the time, no matter how Team 2 plays. Describe such a strategy for Team 1 and explain why it works.

**Problem 18.6.**
Suppose you have a biased coin that has probability $p$ of flipping heads. Let $J$ be the number of heads in $n$ independent coin flips. So $J$ has the general binomial distribution:

$$\mathrm{PDF}_J(k) = \binom{n}{k} p^k q^{n-k}$$

where $q ::= 1 - p$.

**(a)** Show that

$$\begin{aligned} \mathrm{PDF}_J(k-1) &< \mathrm{PDF}_J(k) && \text{for } k < np + p, \\ \mathrm{PDF}_J(k-1) &> \mathrm{PDF}_J(k) && \text{for } k > np + p. \end{aligned}$$

**(b)** Conclude that the maximum value of $\mathrm{PDF}_J$ is asymptotically equal to

$$\frac{1}{\sqrt{2\pi npq}}.$$

*Hint:* For the asymptotic estimate, it's ok to assume that $np$ is an integer, so by part (a), the maximum value is $\mathrm{PDF}_J(np)$. Use Stirling's formula (14.30).

## Homework Problems

**Problem 18.7.**
A drunken sailor wanders along main street, which conveniently consists of the points along the $x$ axis with integral coordinates. In each step, the sailor moves one unit left or right along the $x$ axis. A particular *path* taken by the sailor can be described by a sequence of "left" and "right" steps. For example, ⟨left,left,right⟩ describes the walk that goes left twice then goes right.

We model this scenario with a random walk graph whose vertices are the integers and with edges going in each direction between consecutive integers. All edges are labelled $1/2$.

The sailor begins his random walk at the origin. This is described by an initial distribution which labels the origin with probability 1 and all other vertices with probability 0. After one step, the sailor is equally likely to be at location 1 or $-1$, so the distribution after one step gives label 1/2 to the vertices 1 and $-1$ and labels all other vertices with probability 0.

**(a)** Give the distributions after the 2nd, 3rd, and 4th step by filling in the table of probabilities below, where omitted entries are 0. For each row, write all the nonzero entries so they have the same denominator.

| | location | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | -4 | -3 | -2 | -1 | 0 | 1 | 2 | 3 | 4 |
| initially | | | | | 1 | | | | |
| after 1 step | | | | 1/2 | 0 | 1/2 | | | |
| after 2 steps | | | ? | ? | ? | ? | ? | | |
| after 3 steps | | ? | ? | ? | ? | ? | ? | ? | |
| after 4 steps | ? | ? | ? | ? | ? | ? | ? | ? | ? |

**(b)**

1. What is the final location of a $t$-step path that moves right exactly $i$ times?
2. How many different paths are there that end at that location?
3. What is the probability that the sailor ends at this location?

**(c)** Let $L$ be the random variable giving the sailor's location after $t$ steps, and let $B ::= (L+t)/2$. Use the answer to part (b) to show that $B$ has an unbiased binomial density function.

**(d)** Again let $L$ be the random variable giving the sailor's location after $t$ steps, where $t$ is even. Show that

$$\Pr[|L| < \frac{\sqrt{t}}{2}] < \frac{1}{2}\,.$$

So there is a better than even chance that the sailor ends up at least $\sqrt{t}/2$ steps from where he started.

*Hint:* Work in terms of $B$. Then you can use an estimate that bounds the binomial distribution. Alternatively, observe that the origin is the most likely final location and then use the asymptotic estimate

$$\Pr[L = 0] = \Pr[B = t/2] \sim \sqrt{\frac{2}{\pi t}}.$$

## Problems for Section 18.5

### Practice Problems

**Problem 18.8.**
The vast majority of people have an above average number of fingers. Which of the following statements accounts for this phenomenon? Explain your reasoning.

1. Most people have a super secret extra bonus finger of which they are unaware.
2. A pedantic minority don't count their thumbs as fingers, while the majority of people do.
3. Polydactyly is rarer than amputation.
4. When you add up the total number of fingers among the world's population and then divide by the size of the population, you get a number less than ten.
5. This follows from Markov's Theorem, since no one has a negative number of fingers.
6. Missing fingers are much more common than extra ones.
7. Missing fingers are at least slightly more common than extra ones.

**Problem 18.9.**
A news article reporting on the departure of a school official from California to Alabama dryly commented that this move would raise the average IQ in both states. Explain.

**Problem 18.10.**
MIT students sometimes delay laundry for a few days. Assume all random values described below are mutually independent.

**(a)** A *busy* student must complete 3 problem sets before doing laundry. Each problem set requires 1 day with probability $2/3$ and 2 days with probability $1/3$. Let $B$ be the number of days a busy student delays laundry. What is $\text{Ex}[B]$?

Example: If the first problem set requires 1 day and the second and third problem sets each require 2 days, then the student delays for $B = 5$ days.

**(b)** A *relaxed* student rolls a fair, 6-sided die in the morning. If he rolls a 1, then he does his laundry immediately (with zero days of delay). Otherwise, he delays for one day and repeats the experiment the following morning. Let $R$ be the number of days a relaxed student delays laundry. What is $\text{Ex}[R]$?

Example: If the student rolls a 2 the first morning, a 5 the second morning, and a 1 the third morning, then he delays for $R = 2$ days.

**(c)** Before doing laundry, an *unlucky* student must recover from illness for a number of days equal to the product of the numbers rolled on two fair, 6-sided dice. Let $U$ be the expected number of days an unlucky student delays laundry. What is $\text{Ex}[U]$?

Example: If the rolls are 5 and 3, then the student delays for $U = 15$ days.

**(d)** A student is *busy* with probability $1/2$, *relaxed* with probability $1/3$, and *unlucky* with probability $1/6$. Let $D$ be the number of days the student delays laundry. What is $\text{Ex}[D]$?

**Problem 18.11.**
Each Math for Computer Science final exam will be graded according to a rigorous procedure:

- With probability $\frac{4}{7}$ the exam is graded by a *TA*,with probability $\frac{2}{7}$ it is graded by a *lecturer*, and with probability $\frac{1}{7}$, it is accidentally dropped behind the radiator and arbitrarily given a score of 84.
- *TAs* score an exam by scoring each problem individually and then taking the sum.
  - There are ten true/false questions worth 2 points each. For each, full credit is given with probability $\frac{3}{4}$, and no credit is given with probability $\frac{1}{4}$.
  - There are four questions worth 15 points each. For each, the score is determined by rolling two fair dice, summing the results, and adding 3.
  - The single 20 point question is awarded either 12 or 18 points with equal probability.
- *Lecturers* score an exam by rolling a fair die twice, multiplying the results, and then adding a "general impression"score.
  - With probability $\frac{4}{10}$, the general impression score is 40.
  - With probability $\frac{3}{10}$, the general impression score is 50.
  - With probability $\frac{3}{10}$, the general impression score is 60.

Assume all random choices during the grading process are independent.

**(a)** What is the expected score on an exam graded by a TA?

**(b)** What is the expected score on an exam graded by a lecturer?

**(c)** What is the expected score on a Math for Computer Science final exam?

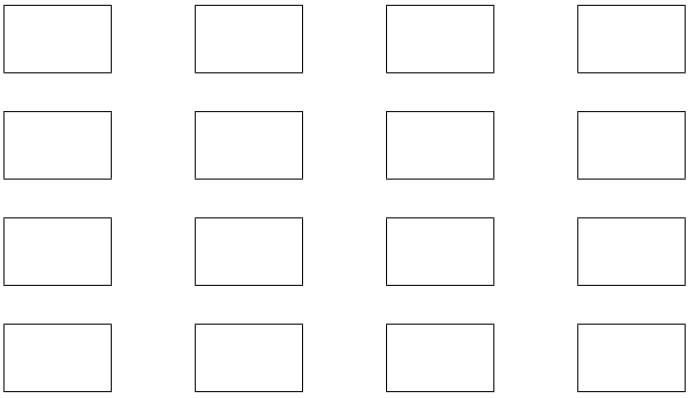
### Class Problems

**Problem 18.12.**
Let's see what it takes to make Carnival Dice fair. Here's the game with payoff parameter $k$: make three independent rolls of a fair die. If you roll a six

- no times, then you lose 1 dollar.
- exactly once, then you win 1 dollar.
- exactly twice, then you win two dollars.
- all three times, then you win $k$ dollars.

For what value of $k$ is this game fair?

**Problem 18.13.**
A classroom has sixteen desks in a $4 \times 4$ arrangement as shown below.



If there is a girl in front, behind, to the left, or to the right of a boy, then the two of them *flirt*. One student may be in multiple flirting couples; for example, a student in a corner of the classroom can flirt with up to two others, while a student in the center can flirt with as many as four others. Suppose that desks are occupied by boys and girls with equal probability and mutually independently. What is the expected number of flirting couples? *Hint:* Linearity.

**Problem 18.14.**
Here are seven propositions:

$$\begin{array}{lllll}
x_1 & \text{OR} & x_3 & \text{OR} & \overline{x_7} \\
\overline{x_5} & \text{OR} & x_6 & \text{OR} & x_7 \\
x_2 & \text{OR} & \overline{x_4} & \text{OR} & x_6 \\
\overline{x_4} & \text{OR} & x_5 & \text{OR} & \overline{x_7} \\
x_3 & \text{OR} & \overline{x_5} & \text{OR} & \overline{x_8} \\
x_9 & \text{OR} & \overline{x_8} & \text{OR} & x_2 \\
\overline{x_3} & \text{OR} & x_9 & \text{OR} & x_4
\end{array}$$

Note that:

1. Each proposition is the disjunction (OR) of three terms of the form $x_i$ or the form $\overline{x_i}$.

2. The variables in the three terms in each proposition are all different.

Suppose that we assign true/false values to the variables $x_1, \ldots, x_9$ independently and with equal probability.

**(a)** What is the expected number of true propositions?

*Hint:* Let $T_i$ be an indicator for the event that the $i$-th proposition is true.

**(b)** Use your answer to prove that for *any* set of 7 propositions satisfying the conditions 1. and 2., there is an assignment to the variables that makes all 7 of the propositions true.

**Problem 18.15.**
A *literal* is a propositional variable or its negation. A *$k$-clause* is an OR of $k$ literals, with no variable occurring more than once in the clause. For example,

$$P \text{ OR } \overline{Q} \text{ OR } \overline{R} \text{ OR } V,$$

is a 4-clause, but

$$\overline{V} \text{ OR } \overline{Q} \text{ OR } \overline{X} \text{ OR } V,$$

is not, since $V$ appears twice.

Let $\mathcal{S}$ be a set of $n$ distinct $k$-clauses involving $v$ variables. The variables in different $k$-clauses may overlap or be completely different, so $k \leq v \leq nk$.

A random assignment of true/false values will be made independently to each of the $v$ variables, with true and false assignments equally likely. Write formulas in $n$, $k$, and $v$ in answer to the first two parts below.

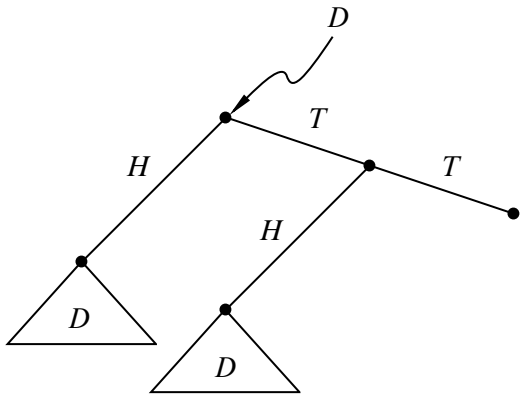



**Figure 18.8** Sample space tree for coin toss until two consective heads.

**(a)** What is the probability that the last $k$-clause in $\mathcal{S}$ is true under the random assignment?

**(b)** What is the expected number of true $k$-clauses in $\mathcal{S}$?

**(c)** A set of propositions is *satisfiable* iff there is an assignment to the variables that makes all of the propositions true. Use your answer to part (b) to prove that if $n < 2^k$, then $\mathcal{S}$ is satisfiable.

**Problem 18.16. (a)** Suppose we flip a fair coin and let $N_{\mathtt{TT}}$ be the number of flips until the first time two Tails in a row appear. What is $\mathrm{Ex}[N_{\mathtt{TT}}]$?

*Hint:* Let $D$ be the tree diagram for this process. Explain why $D$ can be described by the tree in Figure 18.8

Use the **Law of Total Expectation** 18.4.6.

**(b)** Suppose we flip a fair coin until a Tail immediately followed by a Head come up. What is the expectation of the number $N_{\mathtt{TH}}$ of flips we perform?

**(c)** Suppose we now play a game: flip a fair coin until either `TT` or `TH` first occurs. You win if `TT` comes up first, lose if `TH` comes up first. Since `TT` takes 50% longer on average to turn up, your opponent agrees that he has the advantage. So you tell

him you're willing to play if you pay him \$5 when he wins, but he merely pays you a 20% premium, that is, \$6, when you win.

If you do this, you're sneakily taking advantage of your opponent's untrained intuition, since you've gotten him to agree to unfair odds. What is your expected profit per game?

**Problem 18.17.**
Justify each line of the following proof that if $R_1$ and $R_2$ are *independent*, then

$$\mathrm{Ex}[R_1 \cdot R_2] = \mathrm{Ex}[R_1] \cdot \mathrm{Ex}[R_2].$$

*Proof.*

$$\begin{aligned}
&\mathrm{Ex}[R_1 \cdot R_2] \\
&= \sum_{r \in \mathrm{range}(R_1 \cdot R_2)} r \cdot \Pr[R_1 \cdot R_2 = r] \\
&= \sum_{r_i \in \mathrm{range}(R_i)} r_1 r_2 \cdot \Pr[R_1 = r_1 \text{ and } R_2 = r_2] \\
&= \sum_{r_1 \in \mathrm{range}(R_1)} \sum_{r_2 \in \mathrm{range}(R_2)} r_1 r_2 \cdot \Pr[R_1 = r_1 \text{ and } R_2 = r_2] \\
&= \sum_{r_1 \in \mathrm{range}(R_1)} \sum_{r_2 \in \mathrm{range}(R_2)} r_1 r_2 \cdot \Pr[R_1 = r_1] \cdot \Pr[R_2 = r_2] \\
&= \sum_{r_1 \in \mathrm{range}(R_1)} \left( r_1 \Pr[R_1 = r_1] \cdot \sum_{r_2 \in \mathrm{range}(R_2)} r_2 \Pr[R_2 = r_2] \right) \\
&= \sum_{r_1 \in \mathrm{range}(R_1)} r_1 \Pr[R_1 = r_1] \cdot \mathrm{Ex}[R_2] \\
&= \mathrm{Ex}[R_2] \cdot \sum_{r_1 \in \mathrm{range}(R_1)} r_1 \Pr[R_1 = r_1] \\
&= \mathrm{Ex}[R_2] \cdot \mathrm{Ex}[R_1].
\end{aligned}$$

■

**Problem 18.18.**
A gambler bets \$10 on "red" at a roulette table (the odds of red are 18/38 which

slightly less than even) to win \$10. If he wins, he gets back twice the amount of his bet and he quits. Otherwise, he doubles his previous bet and continues.

**(a)** What is the expected number of bets the gambler makes before he wins?

**(b)** What is his probability of winning?

**(c)** What is his expected final profit (amount won minus amount lost)?

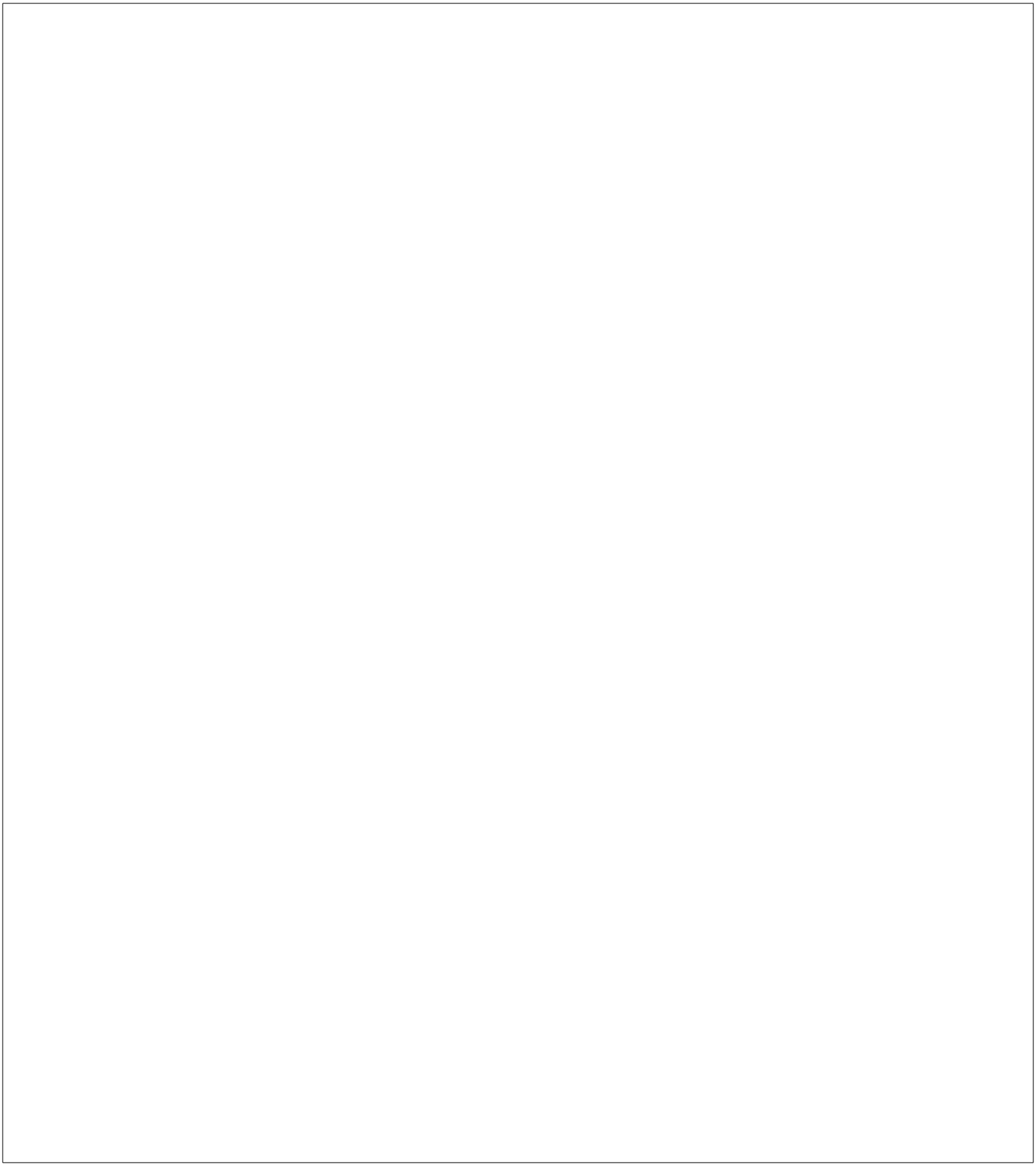
**(d)** The fact that the gambler's expected profit is positive, despite the fact that the game is biased against him, is known as the *St. Petersberg paradox*. The paradox arises from an unrealistic, implicit assumption about the gambler's money. Explain.

*Hint:* What is the expected size of his last bet?

### Homework Problems

**Problem 18.19.**
A coin will be flipped repeatedly until the sequence tail/tail/head (TTH) comes up. Successive flips are independent, and the coin has probability $p$ of coming up heads. Let $N_{\mathtt{TTH}}$ be the number of coin tosses until TTH first appears. What value of $p$ minimizes $\mathrm{Ex}[N_{\mathtt{TTH}}]$?

**Problem 18.20.**
Let $R$ and $S$ be independent random variables, and $f$ and $g$ be any functions such that $\mathrm{domain}(f) = \mathrm{codomain}(R)$ and $\mathrm{domain}(g) = \mathrm{codomain}(S)$. Prove that $f(R)$ and $g(S)$ are independent random variables. *Hint:* The event $[f(R) = a]$ is the disjoint union of all the events $[R = r]$ for $r$ such that $f(r) = a$.



# 19 Deviation from the Mean

## 19.1 Why the Mean?

In the previous chapter we took it for granted that expectation is important, and we developed a bunch of techniques for calculating expected values. But why should we care about this value? After all, a random variable may never take a value anywhere near its expected value.

The most important reason to care about the mean value comes from its connection to estimation by sampling. For example, suppose we want to estimate the average age, income, family size, or other measure of a population. To do this, we determine a random process for selecting people —say throwing darts at census lists. This process makes the selected person's age, income, and so on into a random variable whose *mean* equals the *actual average* age or income of the population. So we can select a random sample of people and calculate the average of people in the sample to estimate the true average in the whole population. But when we make an estimate by repeated sampling, we need to know how much confidence we should have that our estimate is OK or how large a sample is needed to reach a given confidence level. The issue is also fundamental in all experimental science. Because of random errors —*noise* —repeated measurements of the same quantity rarely come out exactly the same. Determining how much confidence to put in experimental measurements is a fundamental and universal scientific issue. Technically, judging sampling or measurement accuracy reduces to finding the probability that an estimate *deviates* by a given amount from its expected value.

Another aspect of this issue comes up in engineering. When designing a sea wall, you need to know how strong to make it to withstand tsunamis for, say, at least a century. If you're assembling a computer network, you need to know how many component failures it should tolerate to likely operate without maintenance for, say, at least a month. If your business is insurance, you need to know how large a financial reserve to maintain to be nearly certain of paying benefits for, say, the next three decades. Technically, such questions come down to finding the probability of *extreme* deviations from the mean.

This issue of *deviation from the mean* is the focus of this chapter.

## 19.2 Markov's Theorem

Markov's theorem gives a generally coarse estimate of the probability that a random variable takes a value *much larger* than its mean. It is an almost trivial result by itself, but it actually leads fairly directly to much stronger results.

The idea behind Markov's Theorem can be explained with a simple example of *intelligence quotient*, IQ. This quantity was devised so that the average IQ measurement would be 100. Now from this fact alone we can conclude that at most 1/3 of the population can have an IQ of 300 or more, because if more than a third had an IQ of 300, then the average would have to be *more* than $(1/3) \cdot 300 = 100$, contradicting the fact that the average is 100. So the probability that a randomly chosen person has an IQ of 300 or more is at most 1/3. Of course this is not a very strong conclusion; in fact no IQ of over 300 has ever been recorded. But by the same logic, we can also conclude that at most 2/3 of the population can have an IQ of 150 or more. IQ's of over 150 have certainly been recorded, though again, a much smaller fraction than 2/3 of the population actually has an IQ that high.

Although these conclusions about IQ are weak, they are actually the strongest general conclusions that can be reached about a random variable using *only* the fact that it is nonnegative and its mean is 100. For example, if we choose a random variable equal to 300 with probability 1/3, and 0 with probability 2/3, then its mean is 100, and the probability of a value of 300 or more really is 1/3. So we can't hope to get a better upper bound based solely on this limited amount of information.

**Theorem 19.2.1** (Markov's Theorem)**.** *If $R$ is a nonnegative random variable, then for all $x > 0$*

$$\Pr[R \geq x] \leq \frac{\mathrm{Ex}[R]}{x}. \tag{19.1}$$

*Proof.* Let $y$ vary over the range of $R$. Then for any $x > 0$

$$\begin{aligned} \mathrm{Ex}[R] ::= & \sum_y y \Pr[R = y] \\ & \geq \sum_{y \geq x} y \Pr[R = y] \geq \sum_{y \geq x} x \Pr[R = y] = x \sum_{y \geq x} \Pr[R = y] \\ & = x \Pr[R \geq x], \end{aligned} \tag{19.2}$$

where the first inequality follows from the fact that $R \geq 0$.

Dividing the first and last expressions in (19.2) by $x$ gives the desired result. ■

Our focus is deviation from the mean, so it's useful to rephrase Markov's Theorem this way:

**Corollary 19.2.2.** *If $R$ is a nonnegative random variable, then for all $c \geq 1$*

$$\Pr[R \geq c \cdot \text{Ex}[R]\,] \leq \frac{1}{c}. \tag{19.3}$$

This Corollary follows immediately from Markov's Theorem(19.2.1) by letting $x$ be $c \cdot \text{Ex}[R]$.

### 19.2.1 Applying Markov's Theorem

Let's go back to the Hat-Check problem of Section 18.5.2. Now we ask what the probability is that $x$ or more men get the right hat, this is, what the value of $\Pr[G \geq x]$ is.

We can compute an upper bound with Markov's Theorem. Since we know $\text{Ex}[G] = 1$, Markov's Theorem implies

$$\Pr[G \geq x] \leq \frac{\text{Ex}[G]}{x} = \frac{1}{x}.$$

For example, there is no better than a 20% chance that 5 men get the right hat, regardless of the number of people at the dinner party.

The Chinese Appetizer problem is similar to the Hat-Check problem. In this case, $n$ people are eating appetizers arranged on a circular, rotating Chinese banquet tray. Someone then spins the tray so that each person receives a random appetizer. What is the probability that everyone gets the same appetizer as before?

There are $n$ equally likely orientations for the tray after it stops spinning. Everyone gets the right appetizer in just one of these $n$ orientations. Therefore, the correct answer is $1/n$.

But what probability do we get from Markov's Theorem? Let the random variable, $R$, be the number of people that get the right appetizer. Then of course $\text{Ex}[R] = 1$ (right?), so applying Markov's Theorem, we find:

$$\Pr[R \geq n] \leq \frac{\text{Ex}[R]}{n} = \frac{1}{n}\,.$$

So for the Chinese appetizer problem, Markov's Theorem is tight!

On the other hand, Markov's Theorem gives the same $1/n$ bound in the Hat-Check problem where the probability of probability everyone gets their hat is $1/(n!)$. So for this case, Markov's Theorem gives a probability bound that is way too large.

### 19.2.2 Markov's Theorem for Bounded Variables

Suppose we learn that the average IQ among MIT students is 150 (which is not true, by the way). What can we say about the probability that an MIT student has an IQ of more than 200? Markov's theorem immediately tells us that no more than $150/200$ or $3/4$ of the students can have such a high IQ. Here we simply applied Markov's Theorem to the random variable, $R$, equal to the IQ of a random MIT student to conclude:

$$\Pr[R > 200] \leq \frac{\mathrm{Ex}[R]}{200} = \frac{150}{200} = \frac{3}{4}.$$

But let's observe an additional fact (which may be true): no MIT student has an IQ less than 100. This means that if we let $T ::= R - 100$, then $T$ is nonnegative and $\mathrm{Ex}[T] = 50$, so we can apply Markov's Theorem to $T$ and conclude:

$$\Pr[R > 200] = \Pr[T > 100] \leq \frac{\mathrm{Ex}[T]}{100} = \frac{50}{100} = \frac{1}{2}.$$

So only half, not 3/4, of the students can be as amazing as they think they are. A bit of a relief!

In fact, we can get better bounds applying Markov's Theorem to $R - b$ instead of $R$ for any lower bound $b > 0$ on $R$ (see Problem 19.2). Similarly, if we have any upper bound, $u$, on a random variable, $S$, then $u - S$ will be a nonnegative random variable, and applying Markov's Theorem to $u - S$ will allow us to bound the probability that $S$ is much *less* than its expectation.

## 19.3 Chebyshev's Theorem

We've seen that Markov's Theorem can give a better bound when applied to $R - b$ rather than $R$. More generally, a good trick for getting stronger bounds on a random variable $R$ out of Markov's Theorem is to apply some cleverly chosen function of $R$.

Choosing functions that are powers of $|R|$ turns out to be specially useful. In particular, since $|R|^\alpha$ is nonnegative, Markov's inequality also applies to the event $[|R|^\alpha \geq x^\alpha]$. But this event is equivalent to the event $[|R| \geq x]$, so we have:

**Lemma 19.3.1.** *For any random variable $R$, $\alpha \in \mathbb{R}^+$, and $x > 0$,*

$$\sigma \Pr[|R| \geq x] \leq \frac{\mathrm{Ex}[|R|^\alpha]}{x^\alpha}.$$

Rephrasing (19.3.1) in terms of the random variable, $|R - \text{Ex}[R]|$, that measures $R$'s deviation from its mean, we get

$$\Pr[\,|R - \text{Ex}[R]\,| \geq x] \leq \frac{\text{Ex}[(R - \text{Ex}[R])^\alpha]}{x^\alpha}. \tag{19.4}$$

The case when $\alpha = 2$ turns out to be so important that the numerator of the right hand side of (19.4) has been given a name:

**Definition 19.3.2.** The *variance*, $\text{Var}[R]$, of a random variable, $R$, is:

$$\text{Var}[R] ::= \text{Ex}\left[(R - \text{Ex}[R])^2\right].$$

The restatement of (19.4) for $\alpha = 2$ is known as *Chebyshev's Theorem*.

**Theorem 19.3.3** (Chebyshev)**.** *Let $R$ be a random variable and $x \in \mathbb{R}^+$. Then*

$$\Pr[|R - \text{Ex}[R]\,| \geq x] \leq \frac{\text{Var}[R]}{x^2}.$$

The expression $\text{Ex}[(R - \text{Ex}[R])^2]$ for variance is a bit cryptic; the best approach is to work through it from the inside out. The innermost expression, $R - \text{Ex}[R]$, is precisely the deviation of $R$ above its mean. Squaring this, we obtain, $(R - \text{Ex}[R])^2$. This is a random variable that is near 0 when $R$ is close to the mean and is a large positive number when $R$ deviates far above or below the mean. So if $R$ is always close to the mean, then the variance will be small. If $R$ is often far from the mean, then the variance will be large.

### 19.3.1 Variance in Two Gambling Games

The relevance of variance is apparent when we compare the following two gambling games.

**Game A:** We win \$2 with probability $2/3$ and lose \$1 with probability $1/3$.

**Game B:** We win \$1002 with probability $2/3$ and lose \$2001 with probability $1/3$.

Which game is better financially? We have the same probability, 2/3, of winning each game, but that does not tell the whole story. What about the expected return for each game? Let random variables $A$ and $B$ be the payoffs for the two games. For example, $A$ is 2 with probability 2/3 and -1 with probability 1/3. We can compute the expected payoff for each game as follows:

$$\text{Ex}[A] = 2 \cdot \frac{2}{3} + (-1) \cdot \frac{1}{3} = 1,$$
$$\text{Ex}[B] = 1002 \cdot \frac{2}{3} + (-2001) \cdot \frac{1}{3} = 1.$$

The expected payoff is the same for both games, but they are obviously very different! This difference is not apparent in their expected value, but is captured by variance. We can compute the Var[$A$] by working "from the inside out" as follows:

$$\begin{aligned} A - \mathrm{Ex}[A] &= \begin{cases} 1 & \text{with probability } \frac{2}{3} \\ -2 & \text{with probability } \frac{1}{3} \end{cases} \\ (A - \mathrm{Ex}[A])^2 &= \begin{cases} 1 & \text{with probability } \frac{2}{3} \\ 4 & \text{with probability } \frac{1}{3} \end{cases} \\ \mathrm{Ex}[(A - \mathrm{Ex}[A])^2] &= 1 \cdot \frac{2}{3} + 4 \cdot \frac{1}{3} \\ \mathrm{Var}[A] &= 2. \end{aligned}$$

Similarly, we have for Var[$B$]:

$$\begin{aligned} B - \mathrm{Ex}[B] &= \begin{cases} 1001 & \text{with probability } \frac{2}{3} \\ -2002 & \text{with probability } \frac{1}{3} \end{cases} \\ (B - \mathrm{Ex}[B])^2 &= \begin{cases} 1,002,001 & \text{with probability } \frac{2}{3} \\ 4,008,004 & \text{with probability } \frac{1}{3} \end{cases} \\ \mathrm{Ex}[(B - \mathrm{Ex}[B])^2] &= 1,002,001 \cdot \frac{2}{3} + 4,008,004 \cdot \frac{1}{3} \\ \mathrm{Var}[B] &= 2,004,002. \end{aligned}$$

The variance of Game A is 2 and the variance of Game B is more than two million! Intuitively, this means that the payoff in Game A is usually close to the expected value of \$1, but the payoff in Game B can deviate very far from this expected value.

High variance is often associated with high risk. For example, in ten rounds of Game A, we expect to make \$10, but could conceivably lose \$10 instead. On the other hand, in ten rounds of game B, we also expect to make \$10, but could actually lose more than \$20,000!

### 19.3.2 Standard Deviation

Because of its definition in terms of the square of a random variable, the variance of a random variable may be very far from a typical deviation from the mean. For example, in Game B above, the deviation from the mean is 1001 in one outcome and -2002 in the other. But the variance is a whopping 2,004,002. From a dimensional analysis viewpoint, the "units" of variance are wrong: if the random variable is in dollars, then the expectation is also in dollars, but the variance is in square dollars. For this reason, people often describe random variables using standard deviation instead of variance.

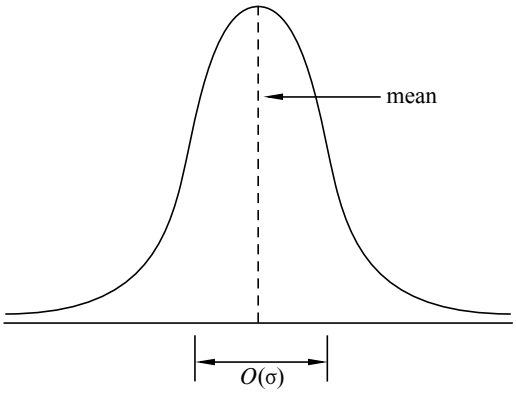



**Figure 19.1** The standard deviation of a distribution indicates how wide the "main part" of it is.

**Definition 19.3.4.** The *standard deviation*, $\sigma_R$, of a random variable, $R$, is the square root of the variance:

$$\sigma_R ::= \sqrt{\text{Var}[R]} = \sqrt{\text{Ex}[(R - \text{Ex}[R])^2]}.$$

So the standard deviation is the square root of the mean of the square of the deviation, or the *root mean square* for short. It has the same units —dollars in our example —as the original random variable and as the mean. Intuitively, it measures the average deviation from the mean, since we can think of the square root on the outside as canceling the square on the inside.

*Example* 19.3.5. The standard deviation of the payoff in Game B is:

$$\sigma_B = \sqrt{\text{Var}[B]} = \sqrt{2,004,002} \approx 1416.$$

The random variable $B$ actually deviates from the mean by either positive 1001 or negative 2002; therefore, the standard deviation of 1416 describes this situation reasonably well.

Intuitively, the standard deviation measures the "width" of the "main part" of the distribution graph, as illustrated in Figure 19.1.

It's useful to rephrase Chebyshev's Theorem in terms of standard deviation which we can do by substituting $x = c\sigma_R$ in (19.1):

**Corollary 19.3.6.** *Let $R$ be a random variable, and let $c$ be a positive real number.*

$$\Pr[|R - \text{Ex}[R]| \geq c\sigma_R] \leq \frac{1}{c^2}. \tag{19.5}$$

Here we see explicitly how the "likely" values of $R$ are clustered in an $O(\sigma_R)$-sized region around $\text{Ex}[R]$, confirming that the standard deviation measures how spread out the distribution of $R$ is around its mean.

**The IQ Example**

Suppose that, in addition to the national average IQ being 100, we also know the standard deviation of IQ's is 10. How rare is an IQ of 300 or more?

Let the random variable, $R$, be the IQ of a random person. So we are supposing that $\text{Ex}[R] = 100$, $\sigma_R = 10$, and $R$ is nonnegative. We want to compute $\Pr[R \geq 300]$.

We have already seen that Markov's Theorem 19.2.1 gives a coarse bound, namely,

$$\Pr[R \geq 300] \leq \frac{1}{3}.$$

Now we apply Chebyshev's Theorem to the same problem:

$$\Pr[R \geq 300] = \Pr[|R - 100| \geq 200] \leq \frac{\text{Var}[R]}{200^2} = \frac{10^2}{200^2} = \frac{1}{400}.$$

So Chebyshev's Theorem implies that at most one person in four hundred has an IQ of 300 or more. We have gotten a much tighter bound using the additional information, namely the variance of $R$, than we could get knowing only the expectation.

## 19.4 Properties of Variance

The definition of variance of $R$ as $\text{Ex}[(R - \text{Ex}[R])^2]$ may seem rather arbitrary. A direct measure of average deviation would be $\text{Ex}[\,|R - \text{Ex}[R]|\,]$. But the direct measure doesn't have the many useful properties that variance has, which is what this section is about.

### 19.4.1 A Formula for Variance

Applying linearity of expectation to the formula for variance yields a convenient alternative formula.

**Lemma 19.4.1.**

$$\text{Var}[R] = \text{Ex}[R^2] - \text{Ex}^2[R],$$

*for any random variable, $R$.*

Here we use the notation $\text{Ex}^2[R]$ as shorthand for $(\text{Ex}[R])^2$.

*Proof.* Let $\mu = \text{Ex}[R]$. Then

$$\begin{aligned}
\text{Var}[R] &= \text{Ex}[(R - \text{Ex}[R])^2] && \text{(Def 19.3.2 of variance)}\\
&= \text{Ex}[(R-\mu)^2] && \text{(def of } \mu\text{)}\\
&= \text{Ex}[R^2 - 2\mu R + \mu^2]\\
&= \text{Ex}[R^2] - 2\mu\,\text{Ex}[R] + \mu^2 && \text{(linearity of expectation)}\\
&= \text{Ex}[R^2] - 2\mu^2 + \mu^2 && \text{(def of } \mu\text{)}\\
&= \text{Ex}[R^2] - \mu^2\\
&= \text{Ex}[R^2] - \text{Ex}^2[R]. && \text{(def of } \mu\text{)}
\end{aligned}$$

■

For example, if $B$ is a Bernoulli variable where $p ::= \Pr[B = 1]$, then

**Lemma 19.4.2.**

$$\text{Var}[B] = p - p^2 = p(1-p). \tag{19.6}$$

*Proof.* By Lemma 18.4.2, $\text{Ex}[B] = p$. But since $B$ only takes values 0 and 1, $B^2 = B$. So Lemma 19.4.2 follows immediately from Lemma 19.4.1. ■

### 19.4.2 Variance of Time to Failure

According to section 18.4.6, the mean time to failure is $1/p$ for a process that fails during any given hour with probability $p$. What about the variance?

By Lemma 19.4.1,

$$\text{Var}[C] = \text{Ex}[C^2] - (1/p)^2 \tag{19.7}$$

so all we need is a formula for $\text{Ex}[C^2]$.

Reasoning about $C$ using conditional expectation worked nicely in section 18.4.6 to find mean time to failure,, and a similar approach works $C^2$. Namely, the expected value of $C^2$ is the probability, $p$, of failure in the first hour times $1^2$, plus the probability, $(1-p)$, of non-failure in the first hour times the expected value of

$(C+1)^2$. So

$$\begin{aligned}
\operatorname{Ex}[C^2] &= p \cdot 1^2 + (1-p)\operatorname{Ex}[(C+1)^2] \\
&= p + (1-p)\left(\operatorname{Ex}[C^2] + \frac{2}{p} + 1\right) \\
&= p + (1-p)\operatorname{Ex}[C^2] + (1-p)\left(\frac{2}{p} + 1\right), \quad \text{so} \\
p\operatorname{Ex}[C^2] &= p + (1-p)\left(\frac{2}{p} + 1\right) \\
&= \frac{p^2 + (1-p)(2+p)}{p} \quad \text{and} \\
\operatorname{Ex}[C^2] &= \frac{2-p}{p^2}
\end{aligned}$$

Combining this with (19.7) proves

**Lemma 19.4.3.** *If failures occur with probability $p$ indendently at each step, and $C$ is the number of steps until the first failure*[1]*, then*

$$\operatorname{Var}[C] = \frac{1-p}{p^2}. \tag{19.8}$$

### 19.4.3 Dealing with Constants

It helps to know how to calculate the variance of $aR + b$:

**Theorem 19.4.4.** *Let $R$ be a random variable, and $a$ a constant. Then*

$$\operatorname{Var}[aR] = a^2 \operatorname{Var}[R]. \tag{19.9}$$

*Proof.* Beginning with the definition of variance and repeatedly applying linearity of expectation, we have:

$$\begin{aligned}
\operatorname{Var}[aR] ::= \operatorname{Ex}[(aR - \operatorname{Ex}[aR])^2] \\
&= \operatorname{Ex}[(aR)^2 - 2aR\operatorname{Ex}[aR] + \operatorname{Ex}^2[aR]] \\
&= \operatorname{Ex}[(aR)^2] - \operatorname{Ex}[2aR\operatorname{Ex}[aR]] + \operatorname{Ex}^2[aR] \\
&= a^2\operatorname{Ex}[R^2] - 2\operatorname{Ex}[aR]\operatorname{Ex}[aR] + \operatorname{Ex}^2[aR] \\
&= a^2\operatorname{Ex}[R^2] - a^2\operatorname{Ex}^2[R] \\
&= a^2\left(\operatorname{Ex}[R^2] - \operatorname{Ex}^2[R]\right) \\
&= a^2\operatorname{Var}[R] \qquad \text{(by Lemma 19.4.1)}
\end{aligned}$$

[1] That is, $C$ has the geometric distribution with parameter $p$ according to Definition 18.4.7.

■

It's even simpler to prove that adding a constant does not change the variance, as the reader can verify:

**Theorem 19.4.5.** *Let $R$ be a random variable, and $b$ a constant. Then*

$$\text{Var}[R + b] = \text{Var}[R]. \tag{19.10}$$

Recalling that the standard deviation is the square root of variance, this implies that the standard deviation of $aR + b$ is simply $|a|$ times the standard deviation of $R$:

**Corollary 19.4.6.**

$$\sigma_{(aR+b)} = |a|\,\sigma_R.$$

### 19.4.4 Variance of a Sum

In general, the variance of a sum is not equal to the sum of the variances, but variances do add for *independent* variables. In fact, *mutual* independence is not necessary: *pairwise* independence will do. This is useful to know because there are some important situations involving variables that are pairwise independent but not mutually independent.

**Theorem 19.4.7.** *If $R_1$ and $R_2$ are independent random variables, then*

$$\text{Var}[R_1 + R_2] = \text{Var}[R_1] + \text{Var}[R_2]. \tag{19.11}$$

*Proof.* We may assume that $\text{Ex}[R_i] = 0$ for $i = 1, 2$, since we could always replace $R_i$ by $R_i - \text{Ex}[R_i]$ in equation (19.11). This substitution preserves the independence of the variables, and by Theorem 19.4.5, does not change the variances.

Now by Lemma 19.4.1, $\text{Var}[R_i] = \text{Ex}[R_i^2]$ and $\text{Var}[R_1 + R_2] = \text{Ex}[(R_1 + R_2)^2]$, so we need only prove

$$\text{Ex}[(R_1 + R_2)^2] = \text{Ex}[R_1^2] + \text{Ex}[R_2^2]. \tag{19.12}$$

But (19.12) follows from linearity of expectation and the fact that

$$\text{Ex}[R_1 R_2] = \text{Ex}[R_1]\,\text{Ex}[R_2] \tag{19.13}$$

since $R_1$ and $R_2$ are independent:

$$\begin{aligned}
\text{Ex}[(R_1 + R_2)^2] &= \text{Ex}[R_1^2 + 2R_1R_2 + R_2^2] \\
&= \text{Ex}[R_1^2] + 2\,\text{Ex}[R_1R_2] + \text{Ex}[R_2^2] \\
&= \text{Ex}[R_1^2] + 2\,\text{Ex}[R_1]\,\text{Ex}[R_2] + \text{Ex}[R_2^2] \qquad \text{(by (19.13))} \\
&= \text{Ex}[R_1^2] + 2 \cdot 0 \cdot 0 + \text{Ex}[R_2^2] \\
&= \text{Ex}[R_1^2] + \text{Ex}[R_2^2]
\end{aligned}$$

■

An independence condition is necessary. If we ignored independence, then we would conclude that $\text{Var}[R+R] = \text{Var}[R]+\text{Var}[R]$. However, by Theorem 19.4.4, the left side is equal to $4\,\text{Var}[R]$, whereas the right side is $2\,\text{Var}[R]$. This implies that $\text{Var}[R] = 0$, which, by the Lemma above, essentially only holds if $R$ is constant.

The proof of Theorem 19.4.7 carries over straightforwardly to the sum of any finite number of variables. So we have:

**Theorem 19.4.8.** *[Pairwise Independent Additivity of Variance] If $R_1, R_2, \ldots, R_n$ are* pairwise *independent random variables, then*

$$\text{Var}[R_1 + R_2 + \cdots + R_n] = \text{Var}[R_1] + \text{Var}[R_2] + \cdots + \text{Var}[R_n]. \tag{19.14}$$

Now we have a simple way of computing the variance of a variable, $J$, that has an $(n, p)$-binomial distribution. We know that $J = \sum_{k=1}^{n} I_k$ where the $I_k$ are mutually independent indicator variables with $\Pr[I_k = 1] = p$. The variance of each $I_k$ is $p(1-p)$ by Lemma 19.4.2, so by linearity of variance, we have

**Lemma** (Variance of the Binomial Distribution)**.** *If $J$ has the $(n, p)$-binomial distribution, then*

$$\text{Var}[J] = n\,\text{Var}[I_k] = np(1-p). \tag{19.15}$$

## 19.5 Estimation by Random Sampling

Democratic politicians were astonished in 2010 when their early polls of sample voters showed Republican Scott Brown was favored by a majority of voters and so would win the special election to fill the Senate seat Democrat Teddy Kennedy had occupied for over 40 years. Based on their poll results, they mounted an intense, but ultimately unsuccessful, effort to save the seat for their party.

### 19.5.1 A Voter Poll

How did polling give an advance estimate of the fraction of the Massachusetts voters who favored Scott Brown over his Democratic opponent?

Suppose at some time before the election that $p$ was the fraction of voters favoring Scott Brown. We want to estimate this unknown fraction $p$. Suppose we have some random process —say throwing darts at voter registration lists —which will select each voter with equal probability. We can define a Bernoulli variable, $K$, by the rule that $K = 1$ if the random voter most prefers Brown, and $K = 0$ otherwise.

Now to estimate $p$, we take a large number, $n$, of random choices of voters[2] and count the fraction who favor Brown. That is, we define variables $K_1, K_2, \ldots$, where $K_i$ is interpreted to be the indicator variable for the event that the $i$th chosen voter prefers Brown. Since our choices are made independently, the $K_i$'s are independent. So formally, we model our estimation process by simply assuming we have mutually independent Bernoulli variables $K_1, K_2, \ldots$, each with the same probability, $p$, of being equal to 1. Now let $S_n$ be their sum, that is,

$$S_n ::= \sum_{i=1}^{n} K_i. \tag{19.16}$$

The variable $S_n/n$ describes the fraction of sampled voters who favor Scott Brown. Most people intuitively expect this sample fraction to give a useful approximation to the unknown fraction, $p$ —and they would be right. So we will use the sample value, $S_n/n$, as our *statistical estimate* of $p$. We know that $S_n$ has the binomial distribution with parameters $n$ and $p$, where we can choose $n$, but $p$ is unknown.

### How Large a Sample?

Suppose we want our estimate to be within 0.04 of the fraction, $p$, at least 95% of the time. This means we want

$$\Pr\left[\left|\frac{S_n}{n} - p\right| \leq 0.04\right] \geq 0.95\,. \tag{19.17}$$

So we better determine the number, $n$, of times we must poll voters so that inequality (19.17) will hold. Chebyshev's Theorem offers a simple way to determine such a $n$.

Since $S_n$ is binomially distributed, equation (19.15) gives

$$\mathrm{Var}[S_n] = n(p(1-p)) \leq n \cdot \frac{1}{4} = \frac{n}{4}.$$

The bound of 1/4 follows from the fact that $p(1-p)$ is maximized when $p = 1-p$, that is, when $p = 1/2$ (check this yourself!).

[2] We're choosing a random voter $n$ times *with replacement*. That is, we don't remove a chosen voter from the set of voters eligible to be chosen later; so we might choose the same voter more than once in $n$ tries! We would get a slightly better estimate if we required $n$ *different* people to be chosen, but doing so complicates both the selection process and its analysis, with little gain in accuracy.

Next, we bound the variance of $S_n/n$:

$$\begin{aligned}
\operatorname{Var}\left[\frac{S_n}{n}\right] &= \left(\frac{1}{n}\right)^2 \operatorname{Var}[S_n] && \text{(by (19.9))} \\
&\le \left(\frac{1}{n}\right)^2 \frac{n}{4} && \text{(by (19.5.1))} \\
&= \frac{1}{4n} && (19.18)
\end{aligned}$$

Using Chebyshev's bound and (19.18) we have:

$$\Pr\left[\left|\frac{S_n}{n} - p\right| \ge 0.04\right] \le \frac{\operatorname{Var}[S_n/n]}{(0.04)^2} = \frac{1}{4n(0.04)^2} = \frac{156.25}{n} \tag{19.19}$$

To make our our estimate with 95% confidence, we want the righthand side of (19.19) to be at most 1/20. So we choose $n$ so that

$$\frac{156.25}{n} \le \frac{1}{20},$$

that is,

$$n \ge 3,125.$$

Section 19.7.2 describes how to get tighter estimates of the tails of binomial distributions that lead to a bound on $n$ that is about four times smaller than the one above. But working through this example using only the variance has the virtue of illustrating an approach to estimation that is applicable to arbitrary random variables, not just binomial variables, and it did lead to a feasible, though larger than necessary, sample size.

### 19.5.2 Matching Birthdays

There are important cases where the relevant distributions are not binomial because the mutual independence properties of the voter preference example do not hold. In these cases, estimation methods based on the Chebyshev bound may be the best approach. Birthday Matching is an example. We already saw in Section 17.6.6 that in a class of 85 students it is virtually certain that two or more students will have the same birthday. This suggests that quite a few pairs of students are likely to have the same birthday. How many?

So as before, suppose there are $n$ students and $d$ days in the year, and let $D$ be the number of pairs of students with the same birthday. Now it will be easy to calculate the expected number of pairs of students with matching birthdays. Then we can

take the same approach as we did in estimating voter preferences to get an estimate of the probability of getting a number of pairs close to the expected number.

Unlike the situation with voter preferences, having matching birthdays for different pairs of students are not mutually independent events, but the matchings are *pairwise independent* —as explained in Section 17.6.6 (and proved in Problem 18.2). This will allow us to apply the same reasoning to Birthday Matching as we did for voter preference. Namely, let $B_1, B_2, \ldots, B_n$ be the birthdays of $n$ independently chosen people, and let $E_{i,j}$ be the indicator variable for the event that the $i$th and $j$th people chosen have the same birthdays, that is, the event $[B_i = B_j]$. So our probability model, the $B_i$'s are mutually independent variables, the $E_{i,j}$'s are pairwise independent. Also, the expectations of $E_{i,j}$ for $i \neq j$ equals the probability that $B_i = B_j$, namely, $1/d$.

Now, $D$, the number of matching pairs of birthdays among the $n$ choices, is simply the sum of the $E_{i,j}$'s:

$$D ::= \sum_{1 \le i < j \le n} E_{i,j}. \tag{19.20}$$

So by linearity of expectation

$$\mathrm{Ex}[D] = \mathrm{Ex}\left[\sum_{1 \le i < j \le n} E_{i,j}\right] = \sum_{1 \le i < j \le n} \mathrm{Ex}[E_{i,j}] = \binom{n}{2} \cdot \frac{1}{d}.$$

Similarly,

$$\begin{aligned}
\mathrm{Var}[D] &= \mathrm{Var}\left[\sum_{1 \le i < j \le n} E_{i,j}\right] \\
&= \sum_{1 \le i < j \le n} \mathrm{Var}[E_{i,j}] && \text{(by Theorem 19.4.8)} \\
&= \binom{n}{2} \cdot \frac{1}{d}\left(1 - \frac{1}{d}\right). && \text{(by Lemma 19.4.2)}
\end{aligned}$$

In particular, for a class of $n = 95$ students with $d = 365$ possible birthdays, we have $\mathrm{Ex}[D] \approx 12.23$ and $\mathrm{Var}[D] \approx 12.23(1 - 1/365) < 12.2$. So by Chebyshev's Theorem

$$\Pr[|D - \mathrm{Ex}[D]| \ge x] < \frac{12.2}{x^2}.$$

Letting $x = 7$, we conclude that there is a better than 75% chance that in a class of 95 students, the number of pairs of students with the same birthday will be within 7 of 12.23, namely will be between 6 and 20.

### 19.5.3 Pairwise Independent Sampling

The reasoning we used above to analyze voter polling and matching birthdays is very similar. We summarize it in slightly more general form with a basic result we call the Pairwise Independent Sampling Theorem. In particular, we do not need to restrict ourselves to sums of zero-one valued variables, or to variables with the same distribution. For simplicity, we state the Theorem for pairwise independent variables with possibly different distributions but with the same mean and variance.

**Theorem 19.5.1** (Pairwise Independent Sampling)**.** *Let $G_1, \ldots, G_n$ be pairwise independent variables with the same mean, $\mu$, and deviation, $\sigma$. Define*

$$S_n ::= \sum_{i=1}^{n} G_i. \tag{19.21}$$

*Then*

$$\Pr\left[\left|\frac{S_n}{n} - \mu\right| \geq x\right] \leq \frac{1}{n}\left(\frac{\sigma}{x}\right)^2.$$

*Proof.* We observe first that the expectation of $S_n/n$ is $\mu$:

$$\begin{aligned}
\operatorname{Ex}\left[\frac{S_n}{n}\right] &= \operatorname{Ex}\left[\frac{\sum_{i=1}^{n} G_i}{n}\right] && \text{(def of } S_n\text{)} \\
&= \frac{\sum_{i=1}^{n} \operatorname{Ex}[G_i]}{n} && \text{(linearity of expectation)} \\
&= \frac{\sum_{i=1}^{n} \mu}{n} \\
&= \frac{n\mu}{n} = \mu.
\end{aligned}$$

The second important property of $S_n/n$ is that its variance is the variance of $G_i$ divided by $n$:

$$\begin{aligned}
\operatorname{Var}\left[\frac{S_n}{n}\right] &= \left(\frac{1}{n}\right)^2 \operatorname{Var}[S_n] && \text{(by (19.9))} \\
&= \frac{1}{n^2} \operatorname{Var}\left[\sum_{i=1}^{n} G_i\right] && \text{(def of } S_n\text{)} \\
&= \frac{1}{n^2} \sum_{i=1}^{n} \operatorname{Var}[G_i] && \text{(pairwise independent additivity)} \\
&= \frac{1}{n^2} \cdot n\sigma^2 = \frac{\sigma^2}{n}. && (19.22)
\end{aligned}$$

This is enough to apply Chebyshev's Theorem and conclude:

$$\begin{aligned}\Pr\left[\left|\frac{S_n}{n}-\mu\right|\geq x\right] &\leq \frac{\mathrm{Var}\left[S_n/n\right]}{x^2}. && \text{(Chebyshev's bound)}\\ &= \frac{\sigma^2/n}{x^2} && \text{(by (19.22))}\\ &= \frac{1}{n}\left(\frac{\sigma}{x}\right)^2.\end{aligned}$$

■

The Pairwise Independent Sampling Theorem provides a precise general statement about how the average of independent samples of a random variable approaches the mean. In particular, it proves what is known as the Law of Large Numbers[3]: by choosing a large enough sample size, we can get arbitrarily accurate estimates of the mean with confidence arbitrarily close to 100%.

**Corollary 19.5.2.** *[Weak Law of Large Numbers] Let $G_1, \ldots, G_n$ be pairwise independent variables with the same mean, $\mu$, and the same finite deviation, and let*

$$S_n ::= \frac{\sum_{i=1}^{n} G_i}{n}.$$

*Then for every $\epsilon > 0$,*

$$\lim_{n\to\infty} \Pr[|S_n - \mu| \leq \epsilon] = 1.$$

## 19.6 Confidence versus Probability

So Chebyshev's Bound implies that sampling 3,125 voters will yield a fraction that, 95% of the time, is within 0.04 of the actual fraction of the voting population who prefer Brown.

Notice that the actual size of the voting population was never considered because *it did not matter*. People who have not studied probability theory often insist that the population size should matter. But our analysis shows that polling a little over 3000 people people is always sufficient, whether there are ten thousand, or a million, or a billion ... voters. You should think about an intuitive explanation that might persuade someone who thinks population size matters.

[3]This is the *Weak* Law of Large Numbers. As you might suppose, there is also a Strong Law, but it's outside the scope of 6.042.

Now suppose a pollster actually takes a sample of 3,125 random voters to estimate the fraction of voters who prefer Brown, and the pollster finds that 1250 of them prefer Brown. It's tempting, **but sloppy**, to say that this means:

**False Claim.** *With probability 0.95, the fraction, $p$, of voters who prefer Brown is* $1250/3125 \pm 0.04$. *Since* $1250/3125 - 0.04 > 1/3$, *there is a 95% chance that more than a third of the voters prefer Brown to all other candidates.*

What's objectionable about this statement is that it talks about the probability or "chance" that a real world fact is true, namely that the actual fraction, $p$, of voters favoring Brown is more than 1/3. But $p$ is what it is, and it simply makes no sense to talk about the probability that it is something else. For example, suppose $p$ is actually 0.3; then it's nonsense to ask about the probability that it is within 0.04 of 1250/3125 —it simply isn't.

This example of voter preference is typical: we want to estimate a fixed, unknown real-world quantity. But *being unknown does not make this quantity a random variable*, so it makes no sense to talk about the probability that it has some property.

A more careful summary of what we have accomplished goes this way:

> We have described a probabilistic procedure for estimating the value of the actual fraction, $p$. The probability that *our estimation procedure* will yield a value within 0.04 of $p$ is 0.95.

This is a bit of a mouthful, so special phrasing closer to the sloppy language is commonly used. The pollster would describe his conclusion by saying that

> At the 95% *confidence level*, the fraction of voters who prefer Brown is $1250/3125 \pm 0.04$.

So confidence levels refer to the results of estimation procedures for real-world quantities. The phrase "confidence level" should be heard as a reminder that some statistical procedure was used to obtain an estimate, and in judging the credibility of the estimate, it may be important to learn just what this procedure was.

## 19.7 Sums of Random Variables

If all you know about a random variable is its mean and variance, then Chebyshev's Theorem is the best you can do when it comes to bounding the probability that the random variable deviates from its mean. In some cases, however, we know

more —for example, that the random variable has a binomial distribution —and then it is possible to prove much stronger bounds. Instead of polynomially small bounds such as $1/c^2$, we can sometimes even obtain exponentially small bounds such as $1/e^c$. As we will soon discover, this is the case whenever the random variable $T$ is the sum of $n$ mutually independent random variables $T_1, T_2, \ldots, T_n$ where $0 \leq T_i \leq 1$. A random variable with a binomial distribution is just one of many examples of such a $T$. Here is another.

### 19.7.1 A Motivating Example

Fussbook is a new social networking site oriented toward unpleasant people.

Like all major web services, Fussbook has a load balancing problem. Specifically, Fussbook receives 24,000 forum posts every 10 minutes. Each post is assigned to one of $m$ computers for processing, and each computer works sequentially through its assigned tasks. Processing an average post takes a computer $1/4$ second. Some posts, such as pointless grammar critiques and snide witticisms, are easier. But the most protracted harangues require 1 full second.

Balancing the work load across the $m$ computers is vital; if any computer is assigned more than 10 minutes of work in a 10-minute interval, then that computer is overloaded and system performance suffers. That would be bad, because Fussbook users are *not* a tolerant bunch.

An early idea was to assign each computer an alphabetic range of forum topics. ("That oughta work!", one programmer said.) But after the computer handling the "*pr*ivacy" and "*pr*eferred text editor" threads melted, the drawback of an ad hoc approach was clear: there are no guarantees.

If the length of every task were known in advance, then finding a balanced distribution would be a kind of "bin packing" problem. Such problems are hard to solve exactly, though approximation algorithms can come close. But in this case, task lengths are not known in advance, which is typical for workload problems in the real world.

So the load balancing problem seems sort of hopeless, because there is no data available to guide decisions. Heck, we might as well assign tasks to computers at random!

As it turns out, random assignment not only balances load reasonably well, but also permits provable performance guarantees in place of "That oughta work!" assertions. In general, a randomized approach to a problem is worth considering when a deterministic solution is hard to compute or requires unavailable information.

Some arithmetic shows that Fussbook's traffic is sufficient to keep $m = 10$ computers running at 100% capacity with perfect load balancing. Surely, more than 10 servers are needed to cope with random fluctuations in task length and imperfect

load balance. But how many is enough? 11? 15? 20? 100? We'll answer that question with a new mathematical tool.

### 19.7.2 The Chernoff Bound

The Chernoff[4] bound is a hammer that you can use to nail a great many problems. Roughly, the Chernoff bound says that certain random variables are very unlikely to significantly exceed their expectation. For example, if the expected load on a computer is just a bit below its capacity, then that computer is unlikely to be overloaded, provided the conditions of the Chernoff bound are satisfied.

More precisely, the Chernoff Bound says that *the sum of lots of little, independent random variables is unlikely to significantly exceed the mean of the sum*. The Markov and Chebyshev bounds lead to the same kind of conclusion but typically provide much weaker bounds. In particular, the Markov and Chebyshev bounds are polynomial, while the Chernoff bound is exponential.

Here is the theorem. The proof will come later in Section 19.7.5.

**Theorem 19.7.1** (Chernoff Bound)**.** *Let $T_1, \ldots T_n$ be mutually independent random variables such that $0 \le T_i \le 1$ for all $i$. Let $T = T_1 + \cdots + T_n$. Then for all $c \ge 1$,*

$$\Pr[T \ge c \operatorname{Ex}[T]] \le e^{-\beta(c)\operatorname{Ex}[T]} \tag{19.23}$$

*where $\beta(c) ::= c \ln c - c + 1$.*

The Chernoff bound applies only to distributions of sums of independent random variables that take on values in the interval $[0, 1]$. The binomial distribution is of course such a distribution, but there are lots of other distributions because the Chernoff bound allows the variables in the sum to have differing, arbitrary, and even unknown distributions over the range $[0, 1]$. Furthermore, there is no direct dependence on the number of random variables in the sum or their expectations. In short, the Chernoff bound gives strong results for lots of problems based on little information —no wonder it is widely used!

### 19.7.3 Chernoff Bound for Binomial Tails

The Chernoff bound is pretty easy to apply, though the details can be daunting at first. Let's walk through a simple example to get the hang of it: getting bounds on the tail of a binomial distribution, for example, bounding the probability that the number of heads that come up in 1000 independent tosses of a coin exceeds the

[4] Yes, this is the same Chernoff who figured out how to beat the state lottery —this guy knows a thing or two.

expectation by 20% or more? Let $T_i$ be an indicator variable for the event that the $i$th coin is heads. Then the total number of heads is

$$T = T_1 + \cdots + T_{1000}.$$

The Chernoff bound requires that the random variables $T_i$ be mutually independent and take on values in the range $[0, 1]$. Both conditions hold here. In this example the $T_i$'s only take the two values 0 and 1, since they're indicators.

The goal is to bound the probability that the number of heads exceeds its expectation by 20% or more; that is, to bound $\Pr[T \geq c\,\mathrm{Ex}[T]]$ where c = 1.2. To that end, we compute $\beta(c)$ as defined in the theorem:

$$\beta(c) = c\ln(c) - c + 1 = 0.0187\ldots.$$

If we assume the coin is fair, then $\mathrm{Ex}[T] = 500$. Plugging these values into the Chernoff bound gives:

$$\begin{aligned}\Pr\left[T \geq 1.2\,\mathrm{Ex}[T]\right] &\leq e^{-\beta(c)\cdot \mathrm{Ex}[T]} \\ &= e^{-(0.0187\ldots)\cdot 500} < 0.0000834.\end{aligned}$$

So the probability of getting 20% or more extra heads on 1000 coins is less than 1 in 10,000.

The bound becomes much stronger as the number of coins increases, because the expected number of heads appears in the exponent of the upper bound. For example, the probability of getting at least 20% extra heads on a million coins is at most

$$e^{-(0.0187\ldots)\cdot 500000} < e^{-9392},$$

which is an inconceivably small number.

Alternatively, the bound also becomes stronger for larger deviations. For example, suppose we're interested in the odds of getting 30% or more extra heads in 1000 tosses, rather than 20%. In that case, $c = 1.3$ instead of 1.2. Consequently, the parameter $\beta(c)$ rises from 0.0187 to about 0.0410, which may not seem significant, but because $\beta(c)$ appears in the exponent of the upper bound, the final probability decreases from around 1 in 10,000 to about 1 in a billion!

### 19.7.4 Chernoff Bound for a Lottery Game

Pick-4 is a lottery game where you pay $1 to pick a 4-digit number between 0000 and 9999. If your number comes up in a random drawing, then you win $5,000. Your chance of winning is 1 in 10,000. If 10 million people play, then the expected number of winners is 1000. When there are exactly 1000 winners, the lottery keeps

$5 million of the $10 million paid for tickets. The lottery operator's nightmare is that the number of winners is much greater —say at the 2000 or greater point where the lottery has to pay out more than it received. What is the probability that will happen?

Let $T_i$ be an indicator for the event that the $i$th player wins. Then $T = T_1 + \cdots + T_n$ is the total number of winners. If we assume[5] that the players' picks and the winning number are random, independent and uniform, then the indicators $T_i$ are independent, as required by the Chernoff bound.

Since 2000 winners would be twice the expected number, we choose $c = 2$, compute $\beta(c) = 0.386\ldots$, and plug these values into the Chernoff bound:

$$\begin{aligned}\Pr[T \geq 2000] &= \Pr\left[T \geq 2\,\mathrm{Ex}[T]\right] \\ &\leq e^{-k\,\mathrm{Ex}[T]} = e^{-(0.386\ldots)\cdot 1000} \\ &< e^{-386}.\end{aligned}$$

So there is almost no chance that the lottery operator pays out double. In fact, the number of winners won't even be 10% higher than expected very often. To prove that, let $c = 1.1$, compute $\beta(c) = 0.00484\ldots$, and plug in again:

$$\begin{aligned}\Pr\left[T \geq 1.1\,\mathrm{Ex}[T]\right] &\leq e^{-k\,\mathrm{Ex}[T]} \\ &= e^{-(0.00484)\cdot 1000} < 0.01.\end{aligned}$$

So the Pick-4 lottery may be exciting for the players, but the lottery operator has little doubt about the outcome!

### Randomized Load Balancing

Now let's return to Fussbook and its load balancing problem. Specifically, we need to determine how many machines suffice to ensure that no server is overloaded; that is, assigned to do more than 10 minutes of work in a 10-minute interval. So a server is overloaded if it gets assigned more than 600 seconds of work.

To begin, let's find the probability that the first server is overloaded. Letting $T$ be the number of seconds of work assigned to the first server, this means we want an upper bound on $\Pr[T \geq 600]$. Let $T_i$ be the number of seconds that the first server spends on the $i$th task: then $T_i$ is zero if the task is assigned to another machine, and otherwise $T_i$ is the length of the task. So $T = \sum_{i=1}^{n} T_i$ is the total length of tasks assigned to the first server, where $n = 24{,}000$.

[5] As we noted in Chapter 18, human choices are often not uniform and they can be highly dependent. For example, lots of people will pick an important date. So the lottery folks should not get too much comfort from the analysis that follows, unless they assign random 4-digit numbers to each player.

The Chernoff bound is applicable only if the $T_i$ are mutually independent and take on values in the range $[0, 1]$. The first condition is satisfied if we assume that task lengths and assignments are independent. And the second condition is satisfied because processing even the most interminable harangue takes at most 1 second.

In all, there are 24,000 tasks, each with an expected length of 1/4 second. Since tasks are assigned to computers at random, the expected load on the first server is:

$$\begin{aligned}\operatorname{Ex}[T] &= \frac{24{,}000 \text{ tasks} \cdot 1/4 \text{ second per task}}{m \text{ machines}} \\ &= 6000/m \text{ seconds.}\end{aligned} \tag{19.24}$$

For example, if there are fewer than 10 machines, then the expected load on the first server is greater than its capacity, and we can expect it to be overloded. If there are exactly 10 machines, then the server is expected to run for $6000/10 = 600$ seconds, which is 100% of its capacity.

Now we can use the Chernoff bound to upper bound the probability that the first server is overloaded. We have from (19.24)

$$600 = c \operatorname{Ex}[T] \qquad \text{where } c ::= m/10,$$

so by the Chernoff bound

$$\Pr[T \geq 600] = \Pr[T \geq c \operatorname{Ex}[T]] \leq e^{-(c \ln(c) - c + 1) \cdot 6000/m},$$

The probability that *some* server is overloaded is at most $m$ times the probability that the first server is overloaded, by the Union Bound in Section 17.4.2. So

$$\begin{aligned}\Pr[\text{some server is overloaded}] &\leq \sum_{i=1}^{m} \Pr[\text{server } i \text{ is overloaded}] \\ &= m \Pr[\text{the first server is overloaded}] \\ &\leq m e^{-(c \ln(c) - c + 1) \cdot 6000/m},\end{aligned}$$

where $c = m/10$. Some values of this upper bound are tabulated below:

$$\begin{aligned} m &= 11: \quad 0.784\ldots \\ m &= 12: \quad 0.000999\ldots \\ m &= 13: \quad 0.0000000760\ldots \end{aligned}$$

These values suggest that a system with $m = 11$ machines might suffer immediate overload, $m = 12$ machines could fail in a few days, but $m = 13$ should be fine for a century or two!

### 19.7.5 Proof of the Chernoff Bound

The proof of the Chernoff bound is somewhat involved. Heck, even *Chernoff* didn't come up with it! His friend, Herman Rubin, showed him the argument. Thinking the bound not very significant, Chernoff did not credit Rubin in print. He felt pretty bad when it became famous![6]

*Proof.* (of Theorem 19.7.1)

For clarity, we'll go through the proof "top down." That is, we'll use facts that are proved immediately afterward.

The key step is to exponentiate both sides of the inequality $T \geq c \operatorname{Ex}[T]$ and then apply the Markov bound:

$$\begin{aligned}
\Pr[T \geq c \operatorname{Ex}[T]] &= \Pr[c^T \geq c^{c \operatorname{Ex}[T]}] \\
&\leq \frac{\operatorname{Ex}[c^T]}{c^{c \operatorname{Ex}[T]}} && \text{(by Markov)} \\
&\leq \frac{e^{(c-1)\operatorname{Ex}[T]}}{c^{c \operatorname{Ex}[T]}} && \text{(by Lemma 19.7.2 below)} \\
&= \frac{e^{(c-1)\operatorname{Ex}[T]}}{e^{c \ln(c) \operatorname{Ex}[T]}} = e^{-(c \ln(c) - c + 1)\operatorname{Ex}[T]}.
\end{aligned}$$

■

Algebra aside, there is a brilliant idea in this proof: in this context, exponentiating somehow supercharges the Markov bound. This is not true in general! One unfortunate side-effect is that we have to bound some nasty expectations involving exponentials in order to complete the proof. This is done in the two lemmas below, where variables take on values as in Theorem 19.7.1.

**Lemma 19.7.2.**

$$\operatorname{Ex}\left[c^T\right] \leq e^{(c-1)\operatorname{Ex}[T]}.$$

[6] See "A Conversation with Herman Chernoff," *Statistical Science* 1996, Vol. 11, No. 4, pp 335–350.

*Proof.*

$$\begin{aligned}
\operatorname{Ex}\left[c^T\right] &= \operatorname{Ex}\left[c^{T_1+\cdots+T_n}\right] && \text{(def of } T\text{)}\\
&= \operatorname{Ex}\left[c^{T_1}\cdots c^{T_n}\right]\\
&= \operatorname{Ex}\left[c^{T_1}\right]\cdots\operatorname{Ex}[c^{T_n}] && \text{(independent product Cor 18.5.7)}\\
&\le e^{(c-1)\operatorname{Ex}[T_1]}\ldots e^{(c-1)\operatorname{Ex}[T_n]} && \text{(by Lemma 19.7.3 below)}\\
&= e^{(c-1)(\operatorname{Ex}[T_1]+\cdots+\operatorname{Ex}[T_n])}\\
&= e^{(c-1)\operatorname{Ex}[T_1+\cdots+T_n]} && \text{(linearity of Ex}[\cdot]\text{)}\\
&= e^{(c-1)\operatorname{Ex}[T]}.
\end{aligned}$$

■

**Lemma 19.7.3.**

$$\operatorname{Ex}[c^{T_i}] \le e^{(c-1)\operatorname{Ex}[T_i]}$$

*Proof.* All summations below range over values $v$ taken by the random variable $T_i$, which are all required to be in the interval $[0, 1]$.

$$\begin{aligned}
\operatorname{Ex}[c^{T_i}] &= \sum c^v \Pr[T_i = v] && \text{(def of Ex}[\cdot]\text{)}\\
&\le \sum (1 + (c-1)v)\Pr[T_i = v] && \text{(convexity —see below)}\\
&= \sum \Pr[T_i = v] + (c-1)v\Pr[T_i = v]\\
&= \sum \Pr[T_i = v] + (c-1)\sum v\Pr[T_i = v]\\
&= 1 + (c-1)\operatorname{Ex}[T_i]\\
&\le e^{(c-1)\operatorname{Ex}[T_i]} && \text{(since } 1 + z \le e^z\text{)}.
\end{aligned}$$

The second step relies on the inequality

$$c^v \le 1 + (c-1)v,$$

which holds for all $v$ in $[0, 1]$ and $c \ge 1$. This follows from the general principle that a convex function, namely $c^v$, is less than the linear function, $1 + (c-1)v$, between their points of intersection, namely $v = 0$ and 1. This inequality is why the variables $T_i$ are restricted to the interval $[0, 1]$. ■

### 19.7.6 Comparing the Bounds

Suppose that we have a collection of mutually independent events $A_1, A_2, \ldots, A_n$, and we want to know how many of the events are likely to occur.

Let $T_i$ be the indicator random variable for $A_i$ and define

$$p_i = \Pr[T_i = 1] = \Pr\left[A_i\right]$$

for $1 \leq i \leq n$. Define

$$T = T_1 + T_2 + \cdots + T_n$$

to be the number of events that occur.

We know from Linearity of Expectation that

$$\begin{aligned} \operatorname{Ex}[T] &= \operatorname{Ex}[T_1] + \operatorname{Ex}[T_2] + \cdots + \operatorname{Ex}[T_n] \\ &= \sum_{i=1}^{n} p_i. \end{aligned}$$

This is true even if the events are *not* independent.

By Theorem 19.4.8, we also know that

$$\begin{aligned} \operatorname{Var}[T] &= \operatorname{Var}[T_1] + \operatorname{Var}[T_2] + \cdots + \operatorname{Var}[T_n] \\ &= \sum_{i=1}^{n} p_i(1 - p_i), \end{aligned}$$

and thus that

$$\sigma_T = \sqrt{\sum_{i=1}^{n} p_i(1 - p_i)}.$$

This is true even if the events are only pairwise independent.

Markov's Theorem tells us that for any $c > 1$,

$$\Pr[T \geq c \operatorname{Ex}[T]] \leq \frac{1}{c}.$$

Chebyshev's Theorem gives us the stronger result that

$$\Pr[|T - \operatorname{Ex}[T]| \geq c\sigma_T] \leq \frac{1}{c^2}.$$

The Chernoff Bound gives us an even stronger result, namely, that for any $c > 0$,

$$\Pr[T - \operatorname{Ex}[T] \geq c \operatorname{Ex}[T]] \leq e^{-(c \ln(c) - c + 1) \operatorname{Ex}[T]}.$$

In this case, the probability of exceeding the mean by $c\,\mathrm{Ex}[T]$ decreases as an exponentially small function of the deviation.

By considering the random variable $n - T$, we can also use the Chernoff Bound to prove that the probability that $T$ is much lower than $\mathrm{Ex}[T]$ is also exponentially small.

### 19.7.7 Murphy's Law

If the expectation of a random variable is much less than 1, then Markov's Theorem implies that there is only a small probability that the variable has a value of 1 or more. On the other hand, a result that we call *Murphy's Law*[7] says that if a random variable is an independent sum of 0-1-valued variables and has a large expectation, then there is a huge probability of getting a value of at least 1.

**Theorem 19.7.4** (Murphy's Law)**.** *Let $A_1, A_2, \ldots, A_n$ be mutually independent events. Let $T_i$ be the indicator random variable for $A_i$ and define*

$$T ::= T_1 + T + 2 + \cdots + T_n$$

*to be the number of events that occur. Then*

$$\Pr[T = 0] \le e^{-\mathrm{Ex}[T]}.$$

*Proof.*

$$\begin{aligned}
\Pr[T = 0] &= \Pr[\overline{A}_1 \wedge \overline{A}_2 \wedge \cdots \wedge \overline{A}_n] \\
&= \prod_{i=1}^{n} \Pr[\overline{A}_i] && \text{(by independence of } A_i\text{)} \\
&= \prod_{i=1}^{n} (1 - \Pr[A_i]) \\
&\le \prod_{i=1}^{n} e^{-\Pr[A_i]} && \text{(since } 1 - x \le e^{-x}\text{)} \\
&= e^{-\sum_{i=1}^{n} \Pr[A_i]} \\
&= e^{-\sum_{i=1}^{n} \mathrm{Ex}[T_i]} && \text{(since } T_i \text{ is an indicator for } A_i\text{)} \\
&= e^{-\mathrm{Ex}[T]} && \text{(linearity of expectation)}
\end{aligned}$$

■

[7] This is in reference and deference to the famous saying that "If something can go wrong, it will go wrong."

For example, given any set of mutually independent events, if you expect 10 of them to happen, then at least one of them will happen with probability at least $1 - e^{-10}$. The probability that none of them happen is at most $e^{-10} < 1/22000$.

So if there are a lot of independent things that can go wrong and their probabilities sum to a number much greater than 1, then Theorem 19.7.4 proves that some of them surely will go wrong.

This result can help to explain "coincidences," "miracles," and crazy events that seem to have been very unlikely to happen. Such events do happen, in part, because there are so many possible unlikely events that the sum of their probabilities is greater than one. For example, someone *does* win the lottery.

In fact, if there are 100,000 random tickets in Pick-4, Theorem 19.7.4 says that the probability that there is no winner is less than $e^{-10} < 1/22000$. More generally, there are literally millions of one-in-a-million possible events and so some of them will surely occur.

## 19.8 Really Great Expectations

Making independent tosses of a fair coin until some desired pattern comes up is a simple process you should feel solidly in command of by now, right? So how about a bet about the simplest such process —tossing until a head comes up? Ok, you're wary of betting with us, but how about this: we'll let *you set the odds*.

### 19.8.1 Repeating Yourself

Here's the bet: you make independent tosses of a fair coin until a head comes up. Then you will repeat the process. If a second head comes up in the same or fewer tosses than the first, you have to start over yet again. You keep starting over until you finally toss a run of tails longer than your first one. The payment rules are that you will pay me 1 cent each time you start over. When you win by finally getting a run of tails longer than your first one, I will pay you some generous amount. And by the way, you're certain to win —whatever your initial run of tails happened to be, a longer run will occur again with probability 1!

For example, if your first tosses are TTTH, then you will keep tossing until you get a run of 4 tails. So your winning flips might be

TTTHTHTTHHTTHTHTTTHTHHHTTTT.

In this run there are 10 heads, which means you had to start over 9 times. So you would have paid me 9 cents by the time you finally won by tossing 4 tails. Now

you've won, and I'll pay you generously —how does 25 cents sound? Maybe you'd rather have \$1? How about \$10?

Of course there's a trap here. Let's calculate your expected winnings.

Suppose your initial run of tails had length $k$. After that, each time a head comes up, you have to start over and try to get $k+1$ tails in a row. If we regard your getting $k+1$ tails in a row as a "failed" try, and regard your having to start over because a head came up too soon as a "successful" try, then the number of times you have to start over is the number of tries till the first failure. So the expected number of tries will be the mean time to failure, which is $2^{k+1}$, because the probability of tossing $k+1$ tails in a row is $2^{-(k+1)}$.

Let $T$ be the length of your initial run of tails. So $T = k$ means that your initial tosses were $\mathtt{T}^k\mathtt{H}$. Let $R$ be the number of times you repeat trying to beat your original run of tails. The number of cents you expect to finish with is the number of cents in my generous payment minus $\mathrm{Ex}[R]$. It's now easy to calculate $\mathrm{Ex}[R]$ by conditioning on the value of $T$:

$$\mathrm{Ex}[R] = \sum_{k\in\mathbb{N}} \mathrm{Ex}[R \mid T = k] \cdot \Pr[T = k] = \sum_{k\in\mathbb{N}} 2^{k+1} \cdot 2^{-(k+1)} = \sum_{k\in\mathbb{N}} 1 = \infty.$$

So you can expect to pay me an infinite number of cents before winning my "generous" payment. No amount of generosity can make this bet fair!

We haven't faced infinite expectations until now, but they just popped up in a very simple way. In fact this particular example is a special case of an astonishingly general one worked out in Problem 19.24: the expected waiting time for *any* random variable to achieve a larger value is infinite.

### 19.8.2 The St. Petersburg Paradox

One of the simplest casino bets is on "red" or "black" at the roulette table. In each play at roulette, a small ball is set spinning around a roulette wheel until it lands in a red, black, or green colored slot. The payoff for a bet on red or black matches the bet; for example, if you bet \$10 on red and the ball lands in a red slot, you get back your original \$10 bet plus another matching \$10.

In the US, a roulette wheel has two green slots among 18 black and 18 red slots, so the probability of red is $18/38 \approx 0.473$. In Europe, where roulette wheels have only one green slot, the odds for red are a little better —that is, $18/37 \approx 0.486$ —but still less than even.

There is a notorious gambling strategy allegedly used against the casino in St. Peterburg way back in czarist days: bet \$10 on red, and keep doubling the bet until a red comes up. This strategy implies that a player will leave the game as a net winner of \$10 as soon as the red first appears.

Suppose you had the good fortune to gamble against a fair roulette wheel. Then whatever your bet on a spin of the wheel, you are equally likely to win or lose, and your expected win is 0. This also means that the expected win after any given number of spins remains zero, so even playing the St. Peterburg strategy it seems your expected win would be 0.

But wait a minute. As long as there is a fixed, positive probability of red appearing on each spin of the wheel, it's *certain* that red will eventually come up. That is, you can be certain of leaving the casino having won \$10. This implies that even against an *unfair* roulette wheel, your expected win is \$10, contradicting the idea that you can't expect to win in a game that's biased against you.

This is paradoxical and something's obviously wrong here. In fact, there are two things wrong.

The first thing that's wrong is the argument claiming that the expectation is 0. It would be 0 if the number of bets had a fixed bound. If you could only make $n$ bets, then your expectation in the fair game would be the sum of your expected wins on each of the bets, namely, $n \cdot 0 = 0$. But there is no such fixed bound, and that changes things.

To explain this carefully, let $C_i$ be the number of dollars won on the $i$th spin. So $C_i = 2^{i-1}$ when red comes up for the first time on the $i$th spin, and $C_i = -2^{i-1}$, when the first red spin comes up after the $i$th spin. We can define $C_i$ to be 0 if the first red comes up before the $i$th spin. This means

$$\mathrm{Ex}[C_i] = 0.$$

Also, the total of your winnings is

$$C ::= \sum_{i \in \mathbb{Z}^+} C_i.$$

The conclusion that $\mathrm{Ex}[C] = 10$ follows from Total Expectation, conditioning on the number of spins till a red first occurs. Namely, if the first red occurs on the $i$th spin, the amount won is

$$-10 \cdot (1 + 2 + 2^2 + \cdots + 2^{i-2}) + 10 \cdot 2^{i-1} = 10.$$

Then by Total Expectation,

$$\begin{aligned} \mathrm{Ex}[C] &= \sum_{i \in \mathbb{Z}^+} \mathrm{Ex}[C \mid \text{first red on } i\text{th spin}] \cdot \Pr[\text{first red on } i\text{th spin}] \\ &= \sum_{i \in \mathbb{Z}^+} 10 \cdot 2^{-i} = 10 \cdot \sum_{i \in \mathbb{Z}^+} 2^{-i} = 10 \cdot 1 = 10. \end{aligned}$$

So sure enough,

$$\mathrm{Ex}[C] ::= \mathrm{Ex}\left[\sum_{i\in\mathbb{Z}^+} C_i\right] = 10. \tag{19.25}$$

But since $\mathrm{Ex}[C_i] = 0$,

$$\sum_{i\in\mathbb{Z}^+} \mathrm{Ex}[C_i] = \sum_{i\in\mathbb{Z}^+} 0 = 0. \tag{19.26}$$

It seems that (19.26) and (19.25) contradict each other, but they don't. The apparent contradiction comes from applying infinite linearity to conclude

**False Claim.**

$$\mathrm{Ex}\left[\sum_{i\in\mathbb{Z}^+} C_i\right] = \sum_{i\in\mathbb{Z}^+} \mathrm{Ex}[C_i].$$

But this is a case where the convergence conditions required for infinite linearity don't hold. Even though the left hand sum converges (to 10) and the right hand sum converges (to 0), the infinite linearity Theorem (18.5.5) requires that the sum of expectations of *absolute values* converges. That is, infinite linearity would follow if the sum

$$\sum_{i\in\mathbb{Z}^+} \mathrm{Ex}[\,|C_i|\,] \tag{19.27}$$

converged. But

$$\begin{aligned}\mathrm{Ex}[\,|C_i|\,] &= (|10\cdot 2^{i-1}|)\cdot \Pr[\text{1st red in } i\text{th spin}]\\ &\quad + (|-10\cdot 2^{i-1}|)\cdot \Pr[\text{1st red after } i\text{th spin}]\\ &\quad + 0\cdot \Pr[\text{1st red before the } i\text{th spin}]\\ &= (10\cdot 2^{i-1})\cdot 2^{-(i)} + (10\cdot 2^{i-1})\cdot 2^{-(i)} + 0 = 10,\end{aligned}$$

so the sum (19.27) diverges —rapidly.

Probability theory truly leads to this absurd conclusion: a game entailing an unbounded number of fair bets may not be fair in the end. In fact, even against an *unfair* wheel, as long as there is some fixed positive probability of red on each spin, you are certain to win \$10 playing the St. Petersburg strategy!

This brings us to the second thing that's wrong here: you may wind up losing a lot of money before you catch up with your net win of \$10. Let $L$ be the number of dollars you need to have in order to keep betting until the wheel finally spins red. If red first comes up on the $i$th spin, then $L$ would equal

$$10(1+2+4+\cdots+2^i) = 10(2^{i+1}-1)$$

By Total Expectation,

$$\begin{aligned}\mathrm{Ex}[L] &= \sum_{i\in\mathbb{Z}^+} \mathrm{Ex}[L \mid \text{1st red in } i\text{th spin}] \cdot \Pr[\text{1st red in } i\text{th spin}] \\ &= \sum_{i\in\mathbb{Z}^+} (10\cdot(2^{i+1}-1))\cdot 2^{-i} \geq \sum_{i\in\mathbb{Z}^+} 10 = \infty.\end{aligned}$$

That is, you can expect to lose an infinite amount of money before finally winning \$10 —giving you a 0% profit.

So yes, probability theory leads to the absurd conclusion that, even with the odds heavily against you, you're certain to win playing roulette, but only if you make the absurd assumption that you have an infinite bankroll. We can't fault the theory for reaching an absurd conclusion from an absurd assumption.

## Problems for Section 19.2

### Class Problems

**Problem 19.1.**
A herd of cows is stricken by an outbreak of *cold cow disease*. The disease lowers the normal body temperature of a cow, and a cow will die if its temperature goes below 90 degrees F. The disease epidemic is so intense that it lowered the average temperature of the herd to 85 degrees. Body temperatures as low as 70 degrees, **but no lower**, were actually found in the herd.

**(a)** Prove that at most 3/4 of the cows could have survived.

*Hint:* Let $T$ be the temperature of a random cow. Make use of Markov's bound.

**(b)** Suppose there are 400 cows in the herd. Show that the bound of part (a) is best possible by giving an example set of temperatures for the cows so that the average herd temperature is 85, and with probability 3/4, a randomly chosen cow will have a high enough temperature to survive.

### Homework Problems

**Problem 19.2.**
If $R$ is a nonnegative random variable, then Markov's Theorem gives an upper bound on $\Pr[R \geq x]$ for any real number $x > \mathrm{Ex}[R]$. If a constant $b \geq 0$ is a lower bound on $R$, then Markov's Theorem can also be applied to $R - b$ to obtain a possibly different bound on $\Pr[R \geq x]$.

**(a)** Show that if $b > 0$, applying Markov's Theorem to $R - b$ gives a smaller upper bound on $\Pr[R \geq x]$ than simply applying Markov's Theorem directly to $R$.

**(b)** What value of $b \geq 0$ in part (a) gives the best bound?

## Problems for Section 19.4

### Practice Problems

**Problem 19.3.**
A gambler plays 120 hands of draw poker, 60 hands of black jack, and 20 hands of stud poker per day. He wins a hand of draw poker with probability 1/6, a hand of black jack with probability 1/2, and a hand of stud poker with probability 1/5.

**(a)** What is the expected number of hands the gambler wins in a day?

**(b)** What would the Markov bound be on the probability that the gambler will win at least 108 hands on a given day?

**(c)** Assume the outcomes of the card games are pairwise independent. What is the variance in the number of hands won per day?

**(d)** What would the Chebyshev bound be on the probability that the gambler will win at least 108 hands on a given day? You may answer with a numerical expression that is not completely evaluated.

**Problem 19.4.** **(a)** A computer program crashes at the end of each hour of use with probability $1/p$, if it has not crashed already. If $H$ is the number of hours until the first crash, we know

$$\mathrm{Ex}[H] = \frac{1}{p}, \qquad \text{(Equation (18.8))}$$

$$\mathrm{Var}[H] = \frac{q}{p^2} \qquad \text{(Equation (19.8)),}$$

where $q ::= 1 - p$.

**(b)** What is the Chebyshev bound on

$$\Pr[|H - (1/p)| > x/p]$$

where $x > 0$?

**(c)** Conclude from part (b) that for $a \geq 2$,

$$\Pr[H > a/p] \leq \frac{1-p}{(a-1)^2}$$

*Hint:* Check that $|H - (1/p)| > (a-1)/p$ iff $H > a/p$.

**(d)** What actually is

$$\Pr[H > a/p]?$$

Conclude that for any fixed $p > 0$, the probability that $H > a/p$ is an asymptotically smaller function of $a$ than the Chebyshev bound of part (c).

### Class Problems

**Problem 19.5.**

The hat-check staff has had a long day serving at a party, and at the end of the party they simply return the $n$ checked hats in a random way such that the probability that any particular person gets their own hat back is $1/n$.

Let $X_i$ be the indicator variable for the $i$th person getting their own hat back. Let $S_n$ be the total number of people who get their own hat back.

**(a)** What is the expected number of people who get their own hat back?

**(b)** Write a simple formula for $\text{Ex}[X_i X_j]$ for $i \neq j$.

*Hint:* What is $\Pr\left[X_j = 1 \mid X_i = 1\right]$?

**(c)** Explain why you cannot use the variance of sums formula to calculate $\text{Var}[S_n]$.

**(d)** Show that $\text{Ex}[S_n^2] = 2$. *Hint:* $X_i^2 = X_i$.

**(e)** What is the variance of $S_n$?

**(f)** Show that there is at most a 1% chance that more than 10 people get their own hat back. Try to give an intuitive explanation of why the chance remains this small regardless of $n$.

**Problem 19.6.**

For any random variable, $R$, with mean, $\mu$, and standard deviation, $\sigma$, the Chebyshev Bound says that for any real number $x > 0$,

$$\Pr[|R - \mu| \geq x] \leq \left(\frac{\sigma}{x}\right)^2.$$

Show that for any real number, $\mu$, and real numbers $x \geq \sigma > 0$, there is an $R$ for which the Chebyshev Bound is tight, that is,

$$\Pr[|R| \geq x] = \left(\frac{\sigma}{x}\right)^2. \tag{19.28}$$

*Hint:* First assume $\mu = 0$ and let $R$ only take values $0, -x$, and $x$.

### Homework Problems

**Problem 19.7.**
There is a "one-sided" version of Chebyshev's bound for deviation above the mean:

**Lemma** (One-sided Chebyshev bound)**.**

$$\Pr[R - \mathrm{Ex}[R] \geq x] \leq \frac{\mathrm{Var}[R]}{x^2 + \mathrm{Var}[R]}.$$

*Hint:* Let $S_a ::= (R - \mathrm{Ex}[R] + a)^2$, for $0 \leq a \in \mathbb{R}$. So $R - \mathrm{Ex}[R] \geq x$ implies $S_a \geq (x + a)^2$. Apply Markov's bound to $\Pr[S_a \geq (x + a)^2]$. Choose $a$ to minimize this last bound.

**Problem 19.8.**
A man has a set of $n$ keys, one of which fits the door to his apartment. He tries the keys until he finds the correct one. Give the expectation and variance for the number of trials until success if

**(a)** he tries the keys at random (possibly repeating a key tried earlier)

**(b)** he chooses keys randomly from among those he has not yet tried.

## Problems for Section 19.6

### Practice Problems

**Problem 19.9.**
You work for the president and you want to estimate the fraction $p$ of voters in the entire nation that will prefer him in the upcoming elections. You do this by random sampling. Specifically, you select a random voter and ask them who they are going to vote for. You do this $n$ times, with each voter selected with uniform probability and independently of other selections. Finally, you use the fraction $P$ of voters who said they will vote for the President as an estimate for $p$.

**(a)** Our theorems about sampling and distributions allow us to calculate how confident we can be that the random variable, $P$, takes a value near the constant, $p$. This calculation uses some facts about voters and the way they are chosen. Circle the true facts among the following:

1. Given a particular voter, the probability of that voter preferring the President is $p$.
2. The probability that some voter is chosen more than once in the random sample goes to one as $n$ increases.

3. The probability that some voter is chosen more than once in the random sample goes to zero as the population of voters grows.
4. All voters are equally likely to be selected as the third in the random sample of $n$ voters (assuming $n \geq 3$).
5. The probability that the second voter in the random sample will favor the President, given that the first voter prefers the President, is greater than $p$.
6. The probability that the second voter in the random sample will favor the President, given that the second voter is from the same state as the first, may not equal $p$.

**(b)** Suppose that according to your calculations, the following is true about your polling:

$$\Pr[|P - p| \leq 0.04] \geq 0.95.$$

You do the asking, you count how many said they will vote for the President, you divide by $n$, and find the fraction is 0.53. Among the following, circle the legitimate things you might say in a call to the President:

1. Mr. President, $p = 0.53$!
2. Mr. President, with probability at least 95 percent, $p$ is within 0.04 of 0.53.
3. Mr. President, either $p$ is within 0.04 of 0.53 or something very strange (5-in-100) has happened.
4. Mr. President, we can be 95% confident that p is within 0.04 of 0.53.

### Class Problems

**Problem 19.10.**

A recent Gallup poll found that 35% of the adult population of the United States believes that the theory of evolution is "well-supported by the evidence." Gallup polled 1928 Americans selected uniformly and independently at random. Of these, 675 asserted belief in evolution, leading to Gallup's estimate that the fraction of Americans who believe in evolution is $675/1928 \approx 0.350$. Gallup claims a margin of error of 3 percentage points, that is, he claims to be confident that his estimate is within 0.03 of the actual percentage.

**(a)** What is the largest variance an indicator variable can have?

**(b)** Use the Pairwise Independent Sampling Theorem to determine a confidence level with which Gallup can make his claim.

**(c)** Gallup actually claims greater than 99% confidence in his estimate. How might he have arrived at this conclusion? (Just explain what quantity he could calculate; you do not need to carry out a calculation.)

**(d)** Accepting the accuracy of all of Gallup's polling data and calculations, can you conclude that there is a high probability that the number of adult Americans who believe in evolution is $35 \pm 3$ percent?

**Problem 19.11.**

Let $B_1, B_2, \ldots, B_n$ be mutually independent random variables with a uniform distribution on the integer interval $[1, d]$. Let $D$ equal to the number of events $[B_i = B_j]$ that happen where $i \neq j$. It was observed in Section 17.6.6 (and proved in Problem 18.2) that $\Pr[B_i = B_j] = 1/d$ for $i \neq j$ and that the events $[B_i = B_j]$ are pairwise independent.

Let $E_{i,j}$ be the indicator variable for the event $[B_i = B_j]$.

**(a)** What are $\mathrm{Ex}[E_{i,j}]$ and $\mathrm{Var}[E_{i,j}]$ for $i \neq j$?

**(b)** What are $\mathrm{Ex}[D]$ and $\mathrm{Var}[D]$?

**(c)** In a 6.01 class of 500 students, the youngest student was born 15 years ago and the oldest 35 years ago. Let $D$ be the number of students in the class who were born on exactly the same date. What is the probability that $4 \leq S \leq 32$? (For simplicity, assume that the distribution of birthdays is uniform over the 7305 days in the two decade interval from 35 years ago to 15 years ago.)

**Problem 19.12.**

A defendent in traffic court is trying to beat a speeding ticket on the grounds that—since virtually everybody speeds on the turnpike—the police have unconstitutional discretion in giving tickets to anyone they choose. (By the way, we don't recommend this defense :-).)

To support his argument, the defendent arranged to get a random sample of trips by 3,125 cars on the turnpike and found that 94% of them broke the speed limit at some point during their trip. He says that as a consequence of sampling theory (in particular, the Pairwise Independent Sampling Theorem), the court can be 95% confident that the actual percentage of all cars that were speeding is $94 \pm 4\%$.

The judge observes that the actual number of car trips on the turnpike was never considered in making this estimate. He is skeptical that, whether there were a thousand, a million, or 100,000,000 car trips on the turnpike, sampling only 3,125

is sufficient to be so confident.

Suppose you were were the defendent. How would you explain to the judge why the number of randomly selected cars that have to be checked for speeding *does not depend on the number of recorded trips*? Remember that judges are not trained to understand formulas, so you have to provide an intuitive, nonquantitative explanation.

**Problem 19.13.**

The proof of the Pairwise Independent Sampling Theorem 19.5.1 was given for a sequence $R_1, R_2, \ldots$ of pairwise independent random variables with the same mean and variance.

The theorem generalizes straighforwardly to sequences of pairwise independent random variables, possibly with *different* distributions, as long as all their variances are bounded by some constant.

**Theorem** (Generalized Pairwise Independent Sampling)**.** *Let $X_1, X_2, \ldots$ be a sequence of pairwise independent random variables such that* $\text{Var}[X_i] \leq b$ *for some* $b \geq 0$ *and all* $i \geq 1$*. Let*

$$A_n ::= \frac{X_1 + X_2 + \cdots + X_n}{n},$$
$$\mu_n ::= \text{Ex}[A_n].$$

*Then for every* $\epsilon > 0$*,*

$$\Pr[|A_n - \mu_n| > \epsilon] \leq \frac{b}{\epsilon^2} \cdot \frac{1}{n}. \tag{19.29}$$

**(a)** Prove the Generalized Pairwise Independent Sampling Theorem.

**(b)** Conclude that the following holds:

**Corollary** (Generalized Weak Law of Large Numbers)**.** *For every* $\epsilon > 0$*,*

$$\lim_{n\to\infty} \Pr[|A_n - \mu_n| \leq \epsilon] = 1.$$

**Problem 19.14.**

An *International Journal of Epidemiology* has a policy of publishing papers about drug trial results only if the conclusion about the drug's effectiveness (or lack thereof) holds at the 95% confidence level. The editors and reviewers carefully check that any trial whose results they publish was *properly performed and accurately reported*. They are also careful to check that trials whose results they publish have been conducted independently of each other.

The editors of the Journal reason that under this policy, their readership can be confident that at most 5% of the published studies will be mistaken. Later, the editors are embarrassed —and astonished —to learn that *every one* of the 20 drug trial results they published during the year was wrong. The editors thought that because the trials were conducted independently, the probability of publishing 20 wrong results was negligible, namely, $(1/20)^{20} < 10^{-25}$.

Write a brief explanation to these befuddled editors explaining what's wrong with their reasoning and how it could be that all 20 published studies were wrong.

*Hint:* xkcd comic: "significant"

## Exam Problems

**Problem 19.15.**

Yesterday, the programmers at a local company wrote a large program. To estimate the fraction, $b$, of lines of code in this program that are buggy, the QA team will take a small sample of lines chosen randomly and independently (so it is possible, though unlikely, that the same line of code might be chosen more than once). For each line chosen, they can run tests that determine whether that line of code is buggy, after which they will use the fraction of buggy lines in their sample as their estimate of the fraction $b$.

The company statistician can use estimates of a binomial distribution to calculate a value, $s$, for a number of lines of code to sample which ensures that with 97% confidence, the fraction of buggy lines in the sample will be within 0.006 of the actual fraction, $b$, of buggy lines in the program.

Mathematically, the *program* is an actual outcome that already happened. The *random sample* is a random variable defined by the process for randomly choosing $s$ lines from the program. The justification for the statistician's confidence depends on some properties of the program and how the random sample of $s$ lines of code from the program are chosen. These properties are described in some of the statements below. Indicate which of these statements are true, and explain your answers.

1. The probability that the ninth line of code in the *program* is buggy is $b$.
2. The probability that the ninth line of code chosen for the *random sample* is defective, is $b$.
3. All lines of code in the program are equally likely to be the third line chosen in the *random sample*.
4. Given that the first line chosen for the *random sample* is buggy, the probability that the second line chosen will also be buggy is greater than $b$.

5. Given that the last line in the *program* is buggy, the probability that the next-to-last line in the program will also be buggy is greater than $b$.

6. The expectation of the indicator variable for the last line in the *random sample* being buggy is $b$.

7. Given that the first two lines of code selected in the *random sample* are the same kind of statement —they might both be assignment statements, or both be conditional statements, or both loop statements,...—the probability that the first line is buggy may be greater than $b$.

8. There is zero probability that all the lines in the *random sample* will be different.

**Problem 19.16.**
Let $G_1, G_2, G_3, \ldots$, be an infinite sequence of pairwise independent random variables with the same expectation, $\mu$, and the same finite variance. Let

$$f(n,\epsilon) ::= \Pr\left[\left|\frac{\sum_{i=1}^{n} G_i}{n} - \mu\right| \leq \epsilon\right].$$

The Weak Law of Large Numbers can be expressed as a logical formula of the form:

$$\forall \epsilon > 0\, Q_1\, Q_2 \ldots [f(n,\epsilon) \geq 1 - \delta]$$

where $Q_1 Q_2 \ldots$ is a sequence of quantifiers from among:

$$\begin{array}{cccccc} \forall n & \exists n & \forall n_0 & \exists n_0 & \forall n \geq n_0 & \exists n \geq n_0 \\ & \forall \delta > 0 & \exists \delta > 0 & \forall \delta \geq 0 & \exists \delta \geq 0 & \end{array}$$

Here the $n$ and $n_0$ range over nonnegative integers, and $\delta$ and $\epsilon$ range over real numbers.

Write out the proper sequence $Q_1 Q_2 \ldots$

## Problems for Section 19.7

### Class Problems

**Problem 19.17.**
We want to store 2 billion records into a hash table that has 1 billion slots. Assuming the records are randomly and independently chosen with uniform probability

of being assigned to each slot, two records are expected to be stored in each slot. Of course under a random assignment, some slots may be assigned more than two records.

**(a)** Show that the probability that a given slot gets assigned more than 23 records is less than $e^{-36}$.

*Hint:* For $c = 12$, the value of $c \ln c - c + 1$ is greater than 18.

**(b)** Show that the probability that there is a slot that gets assigned more than 23 records is less than $e^{-15}$. This is less than $1/3,000,000$. *Hint:* $\ln 10^9 < 21$.

**Problem 19.18.**
Sometimes I forget a few items when I leave the house in the morning. For example, here are probabilities that I forget various pieces of footwear:

| | |
|---|---|
| left sock | 0.2 |
| right sock | 0.1 |
| left shoe | 0.1 |
| right shoe | 0.3 |

**(a)** Let $X$ be the number of these that I forget. What is $\mathrm{Ex}[X]$?

**(b)** Upper bound the probability that I forget one or more items. Make no independence assumptions.

**(c)** Use the Markov Inequality to upper bound the probability that I forget 3 or more items.

**(d)** Now suppose that I forget each item of footwear independently. Use Chebyshev's Inequality to upper bound the probability that I forget two or more items.

**(e)** Use Theorem 19.7.4 to lower bound the probability that I forget one or more items.

**(f)** I'm supposed to remember many other items, of course: clothing, watch, backpack, notebook, pencil, kleenex, ID, keys, etc. Let $X$ be the total number of items I remember. Suppose I remember items mutually independently and $\mathrm{Ex}[X] = 36$. Use Chernoff's Bound to give an upper bound on the probability that I remember 48 or more items.

**(g)** Give an upper bound on the probability that I remember 108 or more items.

**Problem 19.19.**
Reasoning based on the Chernoff bound goes a long way in explaining the recent subprime mortgage collapse. A bit of standard vocabulary about the mortgage market is needed:

- A **loan** is money lent to a borrower. If the borrower does not pay on the loan, the loan is said to be in **default**, and collateral is seized. In the case of mortgage loans, the borrower's home is used as collateral.
- A **bond** is a collection of loans, packaged into one entity. A bond can be divided into **tranches**, in some ordering, which tell us how to assign losses from defaults. Suppose a bond contains 1000 loans, and is divided into 10 tranches of 100 bonds each. Then, all the defaults must fill up the lowest tranche before the affect others. For example, suppose 150 defaults happened. Then, the first 100 defaults would occur in tranche 1, and the next 50 defaults would happen in tranche 2.
- The lowest tranche of a bond is called the **mezzanine tranche**.
- We can make a "super bond" of tranches called a **collateralized debt obligation (CDO)** by collecting mezzanine tranches from different bonds. This super bond can then be itself separated into tranches, which are again ordered to indicate how to assign losses.

**(a)** Suppose that 1000 loans make up a bond, and the fail rate is 5% in a year. Assuming mutual independence, give an upper bound for the probability that there are one or more failures in the second-worst tranche. What is the probability that there are failures in the best Tranche?

**(b)** Now, do not assume that the loans are independent. Give an upper bound for the probability that there are one or more failures in the second tranche. What is an upper bound for the probability that the entire bond defaults? Show that it is a tight bound. *Hint:* Use Markov's theorem.

**(c)** Given this setup (and assuming mutual independence between the loans), what is the expected failure rate in the mezzanine tranche?

**(d)** We take the mezzanine tranches from 100 bonds and create a CDO. What is the expected number of underlying failures to hit the CDO?

**(e)** We divide this CDO into 10 tranches of 1000 bonds each. Assuming mutual independence, give an upper bound on the probability of one or more failures in the best tranche. The third tranche?

**(f)** Repeat the previous question without the assumption of mutual independence.

### Homework Problems

**Problem 19.20.**
An infinite version of Murphy's Law is that if an infinite number of mutually independent events are expected to happen, then the probability that only finitely many happen is 0. This is known as the first *Borel-Cantelli Lemma*.

**(a)** Let $A_0, A_1, \ldots$ be any infinite sequence of mutually independent events such that

$$\sum_{n\in\mathbb{N}} \Pr[A_n] = \infty. \tag{19.30}$$

Prove that $\Pr[\text{no } A_n \text{ occurs}] = 0$.

*Hint:* $B_k$ the event that no $A_n$ with $n \leq k$ occurs. So the event that no $A_n$ occurs is

$$B ::= \bigcap_{k\in\mathbb{N}} B_k.$$

Apply Murphy's Law, Theorem 19.7.4, to $B_k$.

**(b)** Conclude that $\Pr[\text{only finitely many } A_n\text{'s occur}] = 0$.

*Hint:* Let $C_k$ be the event that no $A_n$ with $n \geq k$ occurs. So the event that only finitely many $A_n$'s occur is

$$C ::= \bigcup_{k\in\mathbb{N}} C_k.$$

Apply part (a) to $C_k$.

## Problems for Section 19.8

### Practice Problems

**Problem 19.21.**
Let $R$ be a positive integer valued random variable such that

$$\text{PDF}_R(n) = \frac{1}{cn^3},$$

where

$$c ::= \sum_{n=1}^{\infty} \frac{1}{n^3}.$$

**(a)** Prove that $\text{Ex}[R]$ is finite.

**(b)** Prove that $\text{Var}[R]$ is infinite.

**Problem 19.22.**
Let $T$ be a positive integer valued random variable such that

$$\text{PDF}_T(n) = \frac{1}{an^2},$$

where

$$a ::= \sum_{n \in \mathbb{Z}^+} \frac{1}{n^2}.$$

**(a)** Prove that $\text{Ex}[T]$ is infinite.

**(b)** Prove that $\text{Ex}[\sqrt{T}]$ is finite.

**Class Problems**

**Problem 19.23.**
You have a biased coin with nonzero probability $p < 1$ of coming up heads. You toss until a head comes up, and then, as in Section 19.8, you keep tossing until you get a long run of tails, but this time let "long run" mean a run of tails that is at least $k - 10$ when your initial run was length $k$. Prove that the expected number of times you toss a head and start over is still infinite.

**Problem 19.24.**
Let $T_0, T_1, \ldots$ be a sequence of mutually independent random variables with the same distribution. Let

$$R ::= \min\{k > 0 \mid T_k > T_0\}.$$

**(a)** Suppose the range of the $T_0$ is the set $\{t_0 < t_1 < t_2 < \cdots\}$. Explain why the following Theorem implies that $\text{Ex}[R] = \infty$.

**Theorem 19.8.1.** *If $p_0 + p_1 + p_2 + \cdots = 1$ and all $p_i \geq 0$, then the sum*

$$\Omega ::= \sum_{k \in \mathbb{N}} \frac{p_k}{p_{k+1} + p_{k+2} + \cdots}.$$

*diverges.*

**(b)** Let

$$S_k ::= p_k + p_{k+1} + \ldots,$$

and

$$a_k ::= \frac{S_k}{S_{k+1}} - 1.$$

Prove that

$$\Omega = \sum_{k \in \mathbb{N}} a_k. \tag{19.31}$$

**(c)** Prove that

$$\prod_{k \leq n} (a_k + 1) = \frac{1}{S_{n+1}}.$$

**(d)** Conclude from part (c) that

$$\prod_{k \in \mathbb{N}} (a_k + 1) = \infty. \tag{19.32}$$

**(e)** Conclude that $e^{\Omega} = \infty$ and hence $\Omega = \infty$.

### Exam Problems

**Problem 19.25.**
You have a process for generating a positive integer, $K$. The behavior of your process each time you use it is (mutually) independent of all its other uses. You use your process to generate a random integer, and then use your procedure repeatedly until you generate an integer as big as your first one. Let $R$ be the number of additional integers you have to generate.

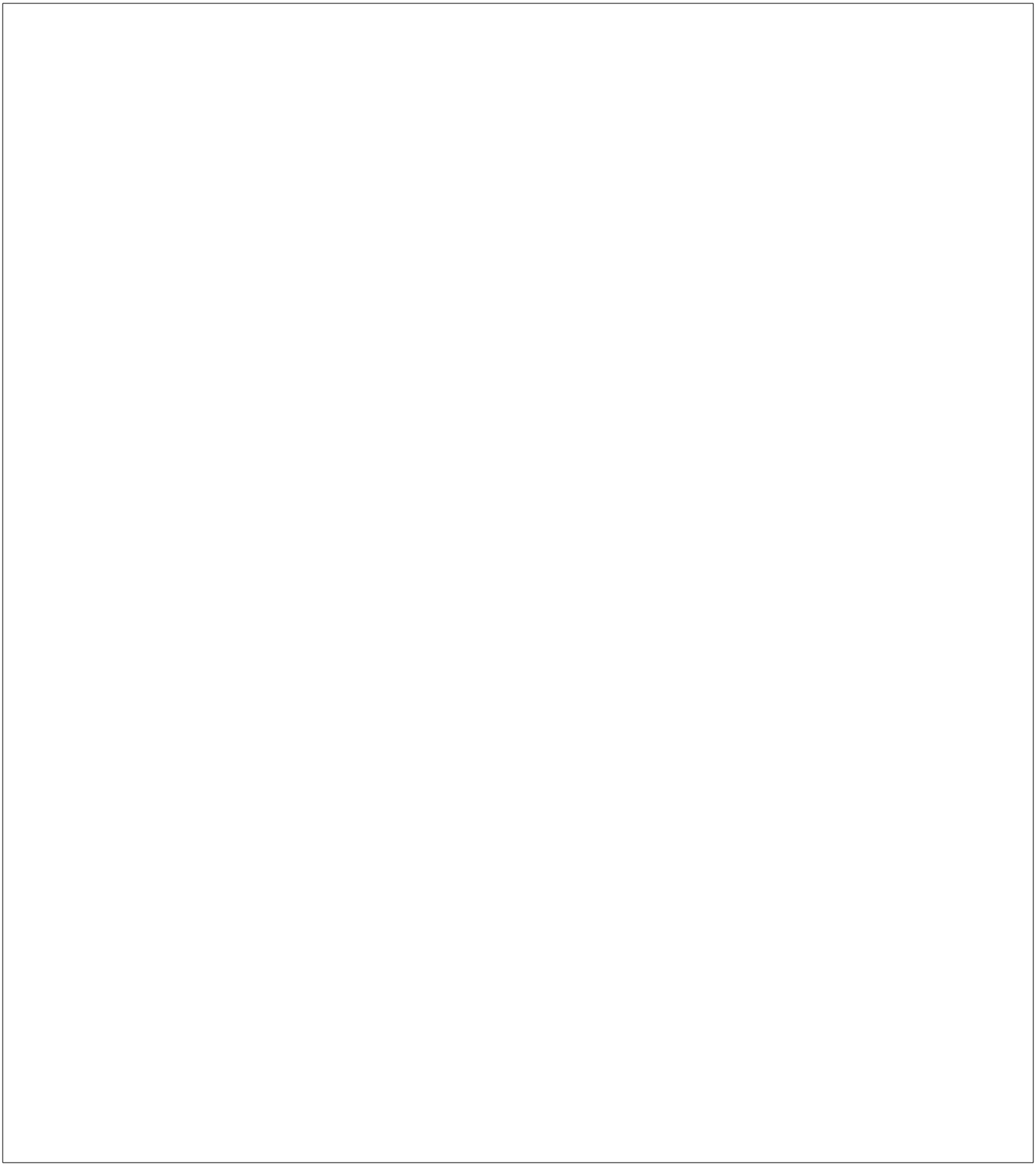
**(a)** State and briefly explain a simple closed formula for $\text{Ex}[R \mid K = k]$ in terms of $\Pr[K \geq k]$.

Suppose $\Pr[K = k] = \Theta(k^{-4})$.

**(b)** Show that $\Pr[K \geq k] = \Theta(k^{-3})$.

**(c)** Show that $\text{Ex}[R]$ is infinite.

# V Recurrences



## Introduction

A *recurrence* describes a sequence of numbers. Early terms are specified explicitly, and later terms are expressed as a function of their predecessors. As a trivial example, here is a recurrence describing the sequence $1, 2, 3, \ldots$:

$$\begin{aligned} T_1 &= 1 \\ T_n &= T_{n-1} + 1 \qquad\qquad (\text{for } n \geq 2). \end{aligned}$$

Here, the first term is defined to be 1 and each subsequent term is one more than its predecessor.

Recurrences turn out to be a powerful tool. In this chapter, we'll emphasize using recurrences to analyze the performance of recursive algorithms. However, recurrences have other applications in computer science as well, such as enumeration of structures and analysis of random processes. And, as we saw in Section 14.4, they also arise in the analysis of problems in the physical sciences.

A recurrence in isolation is not a very useful description of a sequence. One can not easily answer simple questions such as, "What is the hundredth term?" or "What is the asymptotic growth rate?" So one typically wants to *solve* a recurrence; that is, to find a closed-form expression for the $n$th term.

We'll first introduce two general solving techniques: *guess-and-verify* and *plug-and-chug*. These methods are applicable to every recurrence, but their success requires a flash of insight —sometimes an unrealistically brilliant flash. So we'll also introduce two big classes of recurrences, linear and divide-and-conquer, that often come up in computer science. Essentially all recurrences in these two classes are solvable using cookbook techniques; you follow the recipe and get the answer. A drawback is that calculation replaces insight. The "Aha!" moment that is essential in the guess-and-verify and plug-and-chug methods is replaced by a "Huh" at the end of a cookbook procedure.

At the end of the chapter, we'll develop rules of thumb to help you assess many recurrences without any calculation. These rules can help you distinguish promising approaches from bad ideas early in the process of designing an algorithm.

Recurrences are one aspect of a broad theme in computer science: reducing a big problem to progressively smaller problems until easy base cases are reached. This same idea underlies both induction proofs and recursive algorithms. As we'll see, all three ideas snap together nicely. For example, the running time of a recursive algorithm could be described with a recurrence with induction used to verify the solution.

# 21 Recurrences

## 21.1 The Towers of Hanoi

According to legend, there is a temple in Hanoi with three posts and 64 gold disks of different sizes. Each disk has a hole through the center so that it fits on a post. In the misty past, all the disks were on the first post, with the largest on the bottom and the smallest on top, as shown in Figure 21.1.

Monks in the temple have labored through the years since to move all the disks to one of the other two posts according to the following rules:

- The only permitted action is removing the top disk from one post and dropping it onto another post.
- A larger disk can never lie above a smaller disk on any post.

So, for example, picking up the whole stack of disks at once and dropping them on another post is illegal. That's good, because the legend says that when the monks complete the puzzle, the world will end!

To clarify the problem, suppose there were only 3 gold disks instead of 64. Then the puzzle could be solved in 7 steps as shown in Figure 21.2.

The questions we must answer are, "Given sufficient time, can the monks succeed?" If so, "How long until the world ends?" And, most importantly, "Will this happen before the final exam?"

### 21.1.1 A Recursive Solution

The Towers of Hanoi problem can be solved recursively. As we describe the procedure, we'll also analyze the running time. To that end, let $T_n$ be the minimum

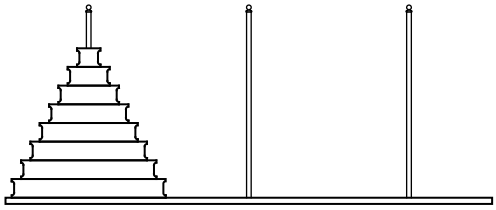


**Figure 21.1** The initial configuration of the disks in the Towers of Hanoi problem.

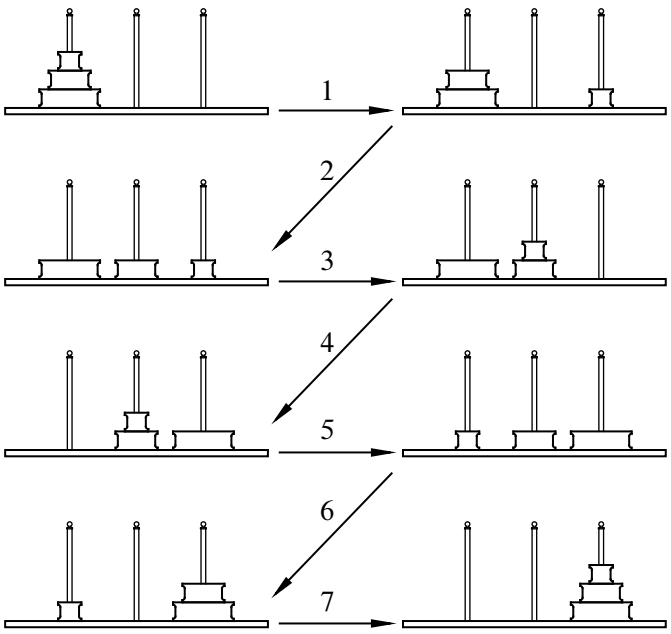



**Figure 21.2** The 7-step solution to the Towers of Hanoi problem when there are $n = 3$ disks.

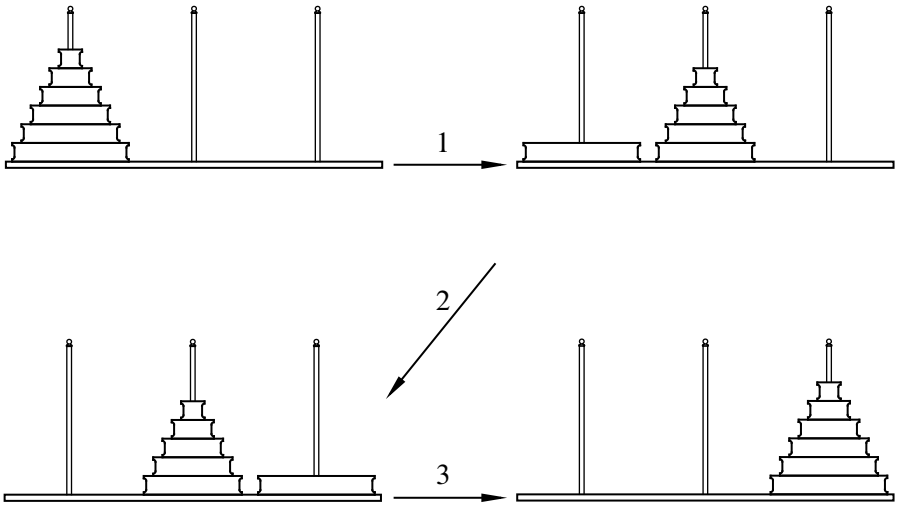



**Figure 21.3** A recursive solution to the Towers of Hanoi problem.

number of steps required to solve the $n$-disk problem. For example, some experimentation shows that $T_1 = 1$ and $T_2 = 3$. The procedure illustrated above shows that $T_3$ is at most 7, though there might be a solution with fewer steps.

The recursive solution has three stages, which are described below and illustrated in Figure 21.3. For clarity, the largest disk is shaded in the figures.

**Stage 1.** Move the top $n-1$ disks from the first post to the second using the solution for $n-1$ disks. This can be done in $T_{n-1}$ steps.

**Stage 2.** Move the largest disk from the first post to the third post. This takes just 1 step.

**Stage 3.** Move the $n-1$ disks from the second post to the third post, again using the solution for $n-1$ disks. This can also be done in $T_{n-1}$ steps.

This algorithm shows that $T_n$, the minimum number of steps required to move $n$ disks to a different post, is at most $T_{n-1} + 1 + T_{n-1} = 2T_{n-1} + 1$. We can use this fact to upper bound the number of operations required to move towers of various heights:

$$T_3 \leq 2 \cdot T_2 + 1 = 7$$
$$T_4 \leq 2 \cdot T_3 + 1 \leq 15$$

Continuing in this way, we could eventually compute an upper bound on $T_{64}$, the number of steps required to move 64 disks. So this algorithm answers our first

question: given sufficient time, the monks can finish their task and end the world. This is a shame. After all that effort, they'd probably want to smack a few high-fives and go out for burgers and ice cream, but nope —world's over.

### 21.1.2 Finding a Recurrence

We can not yet compute the exact number of steps that the monks need to move the 64 disks, only an upper bound. Perhaps, having pondered the problem since the beginning of time, the monks have devised a better algorithm.

In fact, there is no better algorithm, and here is why. At some step, the monks must move the largest disk from the first post to a different post. For this to happen, the $n-1$ smaller disks must all be stacked out of the way on the only remaining post. Arranging the $n-1$ smaller disks this way requires at least $T_{n-1}$ moves. After the largest disk is moved, at least another $T_{n-1}$ moves are required to pile the $n-1$ smaller disks on top.

This argument shows that the number of steps required is at least $2T_{n-1}+1$. Since we gave an algorithm using exactly that number of steps, we can now write an expression for $T_n$, the number of moves required to complete the Towers of Hanoi problem with $n$ disks:

$$\begin{aligned} T_1 &= 1 \\ T_n &= 2T_{n-1}+1 \qquad\qquad (\text{for } n \geq 2). \end{aligned}$$

This is a typical recurrence. These two lines define a sequence of values, $T_1, T_2, T_3, \ldots$. The first line says that the first number in the sequence, $T_1$, is equal to 1. The second line defines every other number in the sequence in terms of its predecessor. So we can use this recurrence to compute any number of terms in the sequence:

$$\begin{aligned} T_1 &= 1 \\ T_2 &= 2\cdot T_1 + 1 = 3 \\ T_3 &= 2\cdot T_2 + 1 = 7 \\ T_4 &= 2\cdot T_3 + 1 = 15 \\ T_5 &= 2\cdot T_4 + 1 = 31 \\ T_6 &= 2\cdot T_5 + 1 = 63. \end{aligned}$$

### 21.1.3 Solving the Recurrence

We could determine the number of steps to move a 64-disk tower by computing $T_7$, $T_8$, and so on up to $T_{64}$. But that would take a lot of work. It would be nice to have a closed-form expression for $T_n$, so that we could quickly find the number of steps required for any given number of disks. (For example, we might want to know how

much sooner the world would end if the monks melted down one disk to purchase burgers and ice cream *before* the end of the world.)

There are several methods for solving recurrence equations. The simplest is to *guess* the solution and then *verify* that the guess is correct with an induction proof. As a basis for a good guess, let's look for a pattern in the values of $T_n$ computed above: 1, 3, 7, 15, 31, 63. A natural guess is $T_n = 2^n - 1$. But whenever you guess a solution to a recurrence, you should always verify it with a proof, typically by induction. After all, your guess might be wrong. (But why bother to verify in this case? After all, if we're wrong, its not the end of the... no, let's check.)

**Claim 21.1.1.** *$T_n = 2^n - 1$ satisfies the recurrence:*

$$\begin{aligned} T_1 &= 1 \\ T_n &= 2T_{n-1} + 1 \qquad\qquad \text{(for } n \geq 2\text{)}. \end{aligned}$$

*Proof.* The proof is by induction on $n$. The induction hypothesis is that $T_n = 2^n - 1$. This is true for $n = 1$ because $T_1 = 1 = 2^1 - 1$. Now assume that $T_{n-1} = 2^{n-1} - 1$ in order to prove that $T_n = 2^n - 1$, where $n \geq 2$:

$$\begin{aligned} T_n &= 2T_{n-1} + 1 \\ &= 2(2^{n-1} - 1) + 1 \\ &= 2^n - 1. \end{aligned}$$

The first equality is the recurrence equation, the second follows from the induction assumption, and the last step is simplification. ■

Such verification proofs are especially tidy because recurrence equations and induction proofs have analogous structures. In particular, the base case relies on the first line of the recurrence, which defines $T_1$. And the inductive step uses the second line of the recurrence, which defines $T_n$ as a function of preceding terms.

Our guess is verified. So we can now resolve our remaining questions about the 64-disk puzzle. Since $T_{64} = 2^{64} - 1$, the monks must complete more than 18 billion billion steps before the world ends. Better study for the final.

### 21.1.4 The Upper Bound Trap

When the solution to a recurrence is complicated, one might try to prove that some simpler expression is an upper bound on the solution. For example, the exact solution to the Towers of Hanoi recurrence is $T_n = 2^n - 1$. Let's try to prove the "nicer" upper bound $T_n \leq 2^n$, proceeding exactly as before.

*Proof.* (Failed attempt.) The proof is by induction on $n$. The induction hypothesis is that $T_n \leq 2^n$. This is true for $n = 1$ because $T_1 = 1 \leq 2^1$. Now assume that $T_{n-1} \leq 2^{n-1}$ in order to prove that $T_n \leq 2^n$, where $n \geq 2$:

$$\begin{aligned} T_n &= 2T_{n-1} + 1 \\ &\leq 2(2^{n-1}) + 1 \\ &\not\leq 2^n \qquad \longleftarrow \text{Uh-oh!} \end{aligned}$$

The first equality is the recurrence relation, the second follows from the induction hypothesis, and the third step is a flaming train wreck. ■

The proof doesn't work! As is so often the case with induction proofs, the argument only goes through with a *stronger* hypothesis. This isn't to say that upper bounding the solution to a recurrence is hopeless, but this is a situation where induction and recurrences do not mix well.

### 21.1.5 Plug and Chug

Guess-and-verify is a simple and general way to solve recurrence equations. But there is one big drawback: you have to *guess right*. That was not hard for the Towers of Hanoi example. But sometimes the solution to a recurrence has a strange form that is quite difficult to guess. Practice helps, of course, but so can some other methods.

Plug-and-chug is another way to solve recurrences. This is also sometimes called "expansion" or "iteration." As in guess-and-verify, the key step is identifying a pattern. But instead of looking at a sequence of *numbers*, you have to spot a pattern in a sequence of *expressions*, which is sometimes easier. The method consists of three steps, which are described below and illustrated with the Towers of Hanoi example.

#### Step 1: Plug and Chug Until a Pattern Appears

The first step is to expand the recurrence equation by alternately "plugging" (applying the recurrence) and "chugging" (simplifying the result) until a pattern appears. Be careful: too much simplification can make a pattern harder to spot. The rule to remember—indeed, a rule applicable to the whole of college life—is *chug in*

*moderation*.

$$\begin{aligned}
T_n &= 2T_{n-1} + 1 \\
&= 2(2T_{n-2} + 1) + 1 && \text{plug} \\
&= 4T_{n-2} + 2 + 1 && \text{chug} \\
&= 4(2T_{n-3} + 1) + 2 + 1 && \text{plug} \\
&= 8T_{n-3} + 4 + 2 + 1 && \text{chug} \\
&= 8(2T_{n-4} + 1) + 4 + 2 + 1 && \text{plug} \\
&= 16T_{n-4} + 8 + 4 + 2 + 1 && \text{chug}
\end{aligned}$$

Above, we started with the recurrence equation. Then we replaced $T_{n-1}$ with $2T_{n-2} + 1$, since the recurrence says the two are equivalent. In the third step, we simplified a little—but not too much! After several similar rounds of plugging and chugging, a pattern is apparent. The following formula seems to hold:

$$\begin{aligned}
T_n &= 2^k T_{n-k} + 2^{k-1} + 2^{k-2} + \cdots + 2^2 + 2^1 + 2^0 \\
&= 2^k T_{n-k} + 2^k - 1
\end{aligned}$$

Once the pattern is clear, simplifying is safe and convenient. In particular, we've collapsed the geometric sum to a closed form on the second line.

**Step 2: Verify the Pattern**

The next step is to verify the general formula with one more round of plug-and-chug.

$$\begin{aligned}
T_n &= 2^k T_{n-k} + 2^k - 1 \\
&= 2^k (2T_{n-(k+1)} + 1) + 2^k - 1 && \text{plug} \\
&= 2^{k+1} T_{n-(k+1)} + 2^{k+1} - 1 && \text{chug}
\end{aligned}$$

The final expression on the right is the same as the expression on the first line, except that $k$ is replaced by $k + 1$. Surprisingly, this effectively *proves* that the formula is correct for all $k$. Here is why: we know the formula holds for $k = 1$, because that's the original recurrence equation. And we've just shown that if the formula holds for some $k \geq 1$, then it also holds for $k + 1$. So the formula holds for all $k \geq 1$ by induction.

**Step 3: Write $T_n$ Using Early Terms with Known Values**

The last step is to express $T_n$ as a function of early terms whose values are known. Here, choosing $k = n - 1$ expresses $T_n$ in terms of $T_1$, which is equal to 1. Simplifying gives a closed-form expression for $T_n$:

$$\begin{aligned} T_n &= 2^{n-1}T_1 + 2^{n-1} - 1 \\ &= 2^{n-1} \cdot 1 + 2^{n-1} - 1 \\ &= 2^n - 1. \end{aligned}$$

We're done! This is the same answer we got from guess-and-verify.

Let's compare guess-and-verify with plug-and-chug. In the guess-and-verify method, we computed several terms at the beginning of the sequence, $T_1$, $T_2$, $T_3$, etc., until a pattern appeared. We generalized to a formula for the $n$th term, $T_n$. In contrast, plug-and-chug works backward from the $n$th term. Specifically, we started with an expression for $T_n$ involving the preceding term, $T_{n-1}$, and rewrote this using progressively earlier terms, $T_{n-2}$, $T_{n-3}$, etc. Eventually, we noticed a pattern, which allowed us to express $T_n$ using the very first term, $T_1$, whose value we knew. Substituting this value gave a closed-form expression for $T_n$. So guess-and-verify and plug-and-chug tackle the problem from opposite directions.

## 21.2 Merge Sort

Algorithms textbooks traditionally claim that sorting is an important, fundamental problem in computer science. Then they smack you with sorting algorithms until life as a disk-stacking monk in Hanoi sounds delightful. Here, we'll cover just *one* well-known sorting algorithm, *Merge Sort*. The analysis introduces another kind of recurrence.

Here is how Merge Sort works. The input is a list of $n$ numbers, and the output is those same numbers in nondecreasing order. There are two cases:

- If the input is a single number, then the algorithm does nothing, because the list is already sorted.
- Otherwise, the list contains two or more numbers. The first half and the second half of the list are each sorted recursively. Then the two halves are merged to form a sorted list with all $n$ numbers.

Let's work through an example. Suppose we want to sort this list:

10, 7, 23, 5, 2, 8, 6, 9.

Since there is more than one number, the first half (10, 7, 23, 5) and the second half (2, 8, 6, 9) are sorted recursively. The results are 5, 7, 10, 23 and 2, 6, 8, 9. All that remains is to merge these two lists. This is done by repeatedly emitting the smaller of the two leading terms. When one list is empty, the whole other list is emitted. The example is worked out below. In this table, underlined numbers are about to be emitted.

| First Half | Second Half | Output |
|---|---|---|
| 5, 7, 10, 23 | 2, 6, 8, 9 | |
| 5, 7, 10, 23 | 6, 8, 9 | 2 |
| 7, 10, 23 | 6, 8, 9 | 2, 5 |
| 7, 10, 23 | 8, 9 | 2, 5, 6 |
| 10, 23 | 8, 9 | 2, 5, 6, 7 |
| 10, 23 | 9 | 2, 5, 6, 7, 8 |
| 10, 23 | | 2, 5, 6, 7, 8, 9 |
| | | 2, 5, 6, 7, 8, 9, 10, 23 |

The leading terms are initially 5 and 2. So we output 2. Then the leading terms are 5 and 6, so we output 5. Eventually, the second list becomes empty. At that point, we output the whole first list, which consists of 10 and 23. The complete output consists of all the numbers in sorted order.

### 21.2.1 Finding a Recurrence

A traditional question about sorting algorithms is, "What is the maximum number of comparisons used in sorting $n$ items?" This is taken as an estimate of the running time. In the case of Merge Sort, we can express this quantity with a recurrence. Let $T_n$ be the maximum number of comparisons used while Merge Sorting a list of $n$ numbers. For now, assume that $n$ is a power of 2. This ensures that the input can be divided in half at every stage of the recursion.

- If there is only one number in the list, then no comparisons are required, so $T_1 = 0$.
- Otherwise, $T_n$ includes comparisons used in sorting the first half (at most $T_{n/2}$), in sorting the second half (also at most $T_{n/2}$), and in merging the two halves. The number of comparisons in the merging step is at most $n - 1$. This is because at least one number is emitted after each comparison and one more number is emitted at the end when one list becomes empty. Since $n$ items are emitted in all, there can be at most $n - 1$ comparisons.

Therefore, the maximum number of comparisons needed to Merge Sort $n$ items is given by this recurrence:

$$\begin{aligned} T_1 &= 0 \\ T_n &= 2T_{n/2} + n - 1 \qquad \text{(for } n \geq 2 \text{ and a power of 2).} \end{aligned}$$

This fully describes the number of comparisons, but not in a very useful way; a closed-form expression would be much more helpful. To get that, we have to solve the recurrence.

### 21.2.2 Solving the Recurrence

Let's first try to solve the Merge Sort recurrence with the guess-and-verify technique. Here are the first few values:

$$\begin{aligned} T_1 &= 0 \\ T_2 &= 2T_1 + 2 - 1 = 1 \\ T_4 &= 2T_2 + 4 - 1 = 5 \\ T_8 &= 2T_4 + 8 - 1 = 17 \\ T_{16} &= 2T_8 + 16 - 1 = 49. \end{aligned}$$

We're in trouble! Guessing the solution to this recurrence is hard because there is no obvious pattern. So let's try the plug-and-chug method instead.

#### Step 1: Plug and Chug Until a Pattern Appears

First, we expand the recurrence equation by alternately plugging and chugging until a pattern appears.

$$\begin{aligned} T_n &= 2T_{n/2} + n - 1 \\ &= 2(2T_{n/4} + n/2 - 1) + (n-1) && \text{plug} \\ &= 4T_{n/4} + (n-2) + (n-1) && \text{chug} \\ &= 4(2T_{n/8} + n/4 - 1) + (n-2) + (n-1) && \text{plug} \\ &= 8T_{n/8} + (n-4) + (n-2) + (n-1) && \text{chug} \\ &= 8(2T_{n/16} + n/8 - 1) + (n-4) + (n-2) + (n-1) && \text{plug} \\ &= 16T_{n/16} + (n-8) + (n-4) + (n-2) + (n-1) && \text{chug} \end{aligned}$$

A pattern is emerging. In particular, this formula seems holds:

$$\begin{aligned} T_n &= 2^k T_{n/2^k} + (n - 2^{k-1}) + (n - 2^{k-2}) + \cdots + (n - 2^0) \\ &= 2^k T_{n/2^k} + kn - 2^{k-1} - 2^{k-2} \cdots - 2^0 \\ &= 2^k T_{n/2^k} + kn - 2^k + 1. \end{aligned}$$

On the second line, we grouped the $n$ terms and powers of 2. On the third, we collapsed the geometric sum.

**Step 2: Verify the Pattern**

Next, we verify the pattern with one additional round of plug-and-chug. If we guessed the wrong pattern, then this is where we'll discover the mistake.

$$\begin{aligned} T_n &= 2^k T_{n/2^k} + kn - 2^k + 1 \\ &= 2^k(2T_{n/2^{k+1}} + n/2^k - 1) + kn - 2^k + 1 && \text{plug} \\ &= 2^{k+1} T_{n/2^{k+1}} + (k+1)n - 2^{k+1} + 1 && \text{chug} \end{aligned}$$

The formula is unchanged except that $k$ is replaced by $k+1$. This amounts to the induction step in a proof that the formula holds for all $k \geq 1$.

**Step 3: Write $T_n$ Using Early Terms with Known Values**

Finally, we express $T_n$ using early terms whose values are known. Specifically, if we let $k = \log n$, then $T_{n/2^k} = T_1$, which we know is 0:

$$\begin{aligned} T_n &= 2^k T_{n/2^k} + kn - 2^k + 1 \\ &= 2^{\log n} T_{n/2^{\log n}} + n\log n - 2^{\log n} + 1 \\ &= nT_1 + n\log n - n + 1 \\ &= n\log n - n + 1. \end{aligned}$$

We're done! We have a closed-form expression for the maximum number of comparisons used in Merge Sorting a list of $n$ numbers. In retrospect, it is easy to see why guess-and-verify failed: this formula is fairly complicated.

As a check, we can confirm that this formula gives the same values that we computed earlier:

| $n$ | $T_n$ | $n\log n - n + 1$ |
|---|---|---|
| 1 | 0 | $1\log 1 - 1 + 1 = 0$ |
| 2 | 1 | $2\log 2 - 2 + 1 = 1$ |
| 4 | 5 | $4\log 4 - 4 + 1 = 5$ |
| 8 | 17 | $8\log 8 - 8 + 1 = 17$ |
| 16 | 49 | $16\log 16 - 16 + 1 = 49$ |

As a double-check, we could write out an explicit induction proof. This would be straightforward, because we already worked out the guts of the proof in step 2 of the plug-and-chug procedure.

## 21.3 Linear Recurrences

So far we've solved recurrences with two techniques: guess-and-verify and plug-and-chug. These methods require spotting a pattern in a sequence of numbers or expressions. In this section and the next, we'll give cookbook solutions for two large classes of recurrences. These methods require no flash of insight; you just follow the recipe and get the answer.

### 21.3.1 Climbing Stairs

How many different ways are there to climb $n$ stairs, if you can either step up one stair or hop up two? For example, there are five different ways to climb four stairs:

1. step, step, step, step
2. hop, hop
3. hop, step, step
4. step, hop step
5. step, step, hop

Working through this problem will demonstrate the major features of our first cookbook method for solving recurrences. We'll fill in the details of the general solution afterward.

#### Finding a Recurrence

As special cases, there is 1 way to climb 0 stairs (do nothing) and 1 way to climb 1 stair (step up). In general, an ascent of $n$ stairs consists of either a step followed by an ascent of the remaining $n-1$ stairs or a hop followed by an ascent of $n-2$ stairs. So the total number of ways to climb $n$ stairs is equal to the number of ways to climb $n-1$ plus the number of ways to climb $n-2$. These observations define a recurrence:

$$\begin{aligned} f(0) &= 1 \\ f(1) &= 1 \\ f(n) &= f(n-1) + f(n-2) \qquad \text{for } n \geq 2. \end{aligned}$$

Here, $f(n)$ denotes the number of ways to climb $n$ stairs. Also, we've switched from subscript notation to functional notation, from $T_n$ to $f_n$. Here the change is cosmetic, but the expressiveness of functions will be useful later.

This is the Fibonacci recurrence, the most famous of all recurrence equations. Fibonacci numbers arise in all sorts of applications and in nature. Fibonacci introduced the numbers in 1202 to study rabbit reproduction. Fibonacci numbers also appear, oddly enough, in the spiral patterns on the faces of sunflowers. And the input numbers that make Euclid's GCD algorithm require the greatest number of steps are consecutive Fibonacci numbers.

### Solving the Recurrence

The Fibonacci recurrence belongs to the class of linear recurrences, which are essentially all solvable with a technique that you can learn in an hour. This is somewhat amazing, since the Fibonacci recurrence remained unsolved for almost six centuries!

In general, a *homogeneous linear recurrence* has the form

$$f(n) = a_1 f(n-1) + a_2 f(n-2) + \cdots + a_d f(n-d)$$

where $a_1, a_2, \ldots, a_d$ and $d$ are constants. The *order* of the recurrence is $d$. Commonly, the value of the function $f$ is also specified at a few points; these are called *boundary conditions*. For example, the Fibonacci recurrence has order $d = 2$ with coefficients $a_1 = a_2 = 1$ and $g(n) = 0$. The boundary conditions are $f(0) = 1$ and $f(1) = 1$. The word "homogeneous" sounds scary, but effectively means "the simpler kind." We'll consider linear recurrences with a more complicated form later.

Let's try to solve the Fibonacci recurrence with the benefit centuries of hindsight. In general, linear recurrences tend to have exponential solutions. So let's guess that

$$f(n) = x^n$$

where $x$ is a parameter introduced to improve our odds of making a correct guess. We'll figure out the best value for $x$ later. To further improve our odds, let's neglect the boundary conditions, $f(0) = 0$ and $f(1) = 1$, for now. Plugging this guess into the recurrence $f(n) = f(n-1) + f(n-2)$ gives

$$x^n = x^{n-1} + x^{n-2}.$$

Dividing both sides by $x^{n-2}$ leaves a quadratic equation:

$$x^2 = x + 1.$$

Solving this equation gives *two* plausible values for the parameter $x$:

$$x = \frac{1 \pm \sqrt{5}}{2}.$$

This suggests that there are at least two different solutions to the recurrence, neglecting the boundary conditions.

$$f(n) = \left(\frac{1+\sqrt{5}}{2}\right)^n \quad \text{or} \quad f(n) = \left(\frac{1-\sqrt{5}}{2}\right)^n$$

A charming features of homogeneous linear recurrences is that any linear combination of solutions is another solution.

**Theorem 21.3.1.** *If $f(n)$ and $g(n)$ are both solutions to a homogeneous linear recurrence, then $h(n) = sf(n) + tg(n)$ is also a solution for all $s, t \in \mathbb{R}$.*

*Proof.*

$$\begin{aligned} h(n) &= sf(n) + tg(n) \\ &= s\left(a_1 f(n-1) + \cdots + a_d f(n-d)\right) + t\left(a_1 g(n-1) + \cdots + a_d g(n-d)\right) \\ &= a_1(sf(n-1) + tg(n-1)) + \cdots + a_d(sf(n-d) + tg(n-d)) \\ &= a_1 h(n-1) + \cdots + a_d h(n-d) \end{aligned}$$

The first step uses the definition of the function $h$, and the second uses the fact that $f$ and $g$ are solutions to the recurrence. In the last two steps, we rearrange terms and use the definition of $h$ again. Since the first expression is equal to the last, $h$ is also a solution to the recurrence. ■

The phenomenon described in this theorem —a linear combination of solutions is another solution —also holds for many differential equations and physical systems. In fact, linear recurrences are so similar to linear differential equations that you can safely snooze through that topic in some future math class.

Returning to the Fibonacci recurrence, this theorem implies that

$$f(n) = s\left(\frac{1+\sqrt{5}}{2}\right)^n + t\left(\frac{1-\sqrt{5}}{2}\right)^n$$

is a solution for all real numbers $s$ and $t$. The theorem expanded two solutions to a whole spectrum of possibilities! Now, given all these options to choose from, we can find one solution that satisfies the boundary conditions, $f(0) = 1$ and $f(1) = 1$. Each boundary condition puts some constraints on the parameters $s$ and $t$. In particular, the first boundary condition implies that

$$f(0) = s\left(\frac{1+\sqrt{5}}{2}\right)^0 + t\left(\frac{1-\sqrt{5}}{2}\right)^0 = s + t = 1.$$

Similarly, the second boundary condition implies that

$$f(1) = s\left(\frac{1+\sqrt{5}}{2}\right)^1 + t\left(\frac{1-\sqrt{5}}{2}\right)^1 = 1.$$

Now we have two linear equations in two unknowns. The system is not degenerate, so there is a unique solution:

$$s = \frac{1}{\sqrt{5}} \cdot \frac{1+\sqrt{5}}{2} \qquad t = -\frac{1}{\sqrt{5}} \cdot \frac{1-\sqrt{5}}{2}.$$

These values of $s$ and $t$ identify a solution to the Fibonacci recurrence that also satisfies the boundary conditions:

$$\begin{aligned} f(n) &= \frac{1}{\sqrt{5}} \cdot \frac{1+\sqrt{5}}{2}\left(\frac{1+\sqrt{5}}{2}\right)^n - \frac{1}{\sqrt{5}} \cdot \frac{1-\sqrt{5}}{2}\left(\frac{1-\sqrt{5}}{2}\right)^n \\ &= \frac{1}{\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)^{n+1} - \frac{1}{\sqrt{5}}\left(\frac{1-\sqrt{5}}{2}\right)^{n+1}. \end{aligned}$$

It is easy to see why no one stumbled across this solution for almost six centuries. All Fibonacci numbers are integers, but this expression is full of square roots of five! Amazingly, the square roots always cancel out. This expression really does give the Fibonacci numbers if we plug in $n = 0, 1, 2$, etc.

This closed-form for Fibonacci numbers has some interesting corollaries. The first term tends to infinity because the base of the exponential, $(1+\sqrt{5})/2 = 1.618\ldots$ is greater than one. This value is often denoted $\phi$ and called the "golden ratio." The second term tends to zero, because $(1-\sqrt{5})/2 = -0.618033988\ldots$ has absolute value less than 1. This implies that the $n$th Fibonacci number is:

$$f(n) = \frac{\phi^{n+1}}{\sqrt{5}} + o(1).$$

Remarkably, this expression involving irrational numbers is actually very close to an integer for all large $n$ —namely, a Fibonacci number! For example:

$$\frac{\phi^{20}}{\sqrt{5}} = 6765.000029\cdots \approx f(19).$$

This also implies that the ratio of consecutive Fibonacci numbers rapidly approaches the golden ratio. For example:

$$\frac{f(20)}{f(19)} = \frac{10946}{6765} = 1.618033998\ldots.$$

### 21.3.2 Solving Homogeneous Linear Recurrences

The method we used to solve the Fibonacci recurrence can be extended to solve any homogeneous linear recurrence; that is, a recurrence of the form

$$f(n) = a_1 f(n-1) + a_2 f(n-2) + \cdots + a_d f(n-d)$$

where $a_1, a_2, \ldots, a_d$ and $d$ are constants. Substituting the guess $f(n) = x^n$, as with the Fibonacci recurrence, gives

$$x^n = a_1 x^{n-1} + a_2 x^{n-2} + \cdots + a_d x^{n-d}.$$

Dividing by $x^{n-d}$ gives

$$x^d = a_1 x^{d-1} + a_2 x^{d-2} + \cdots + a_{d-1} x + a_d.$$

This is called the *characteristic equation* of the recurrence. The characteristic equation can be read off quickly since the coefficients of the equation are the same as the coefficients of the recurrence.

The solutions to a linear recurrence are defined by the roots of the characteristic equation. Neglecting boundary conditions for the moment:

- If $r$ is a nonrepeated root of the characteristic equation, then $r^n$ is a solution to the recurrence.
- If $r$ is a repeated root with multiplicity $k$ then $r^n$, $nr^n$, $n^2 r^n$, ..., $n^{k-1} r^n$ are all solutions to the recurrence.

Theorem 21.3.1 implies that every linear combination of these solutions is also a solution.

For example, suppose that the characteristic equation of a recurrence has roots $s$, $t$, and $u$ twice. These four roots imply four distinct solutions:

$$f(n) = s^n \qquad f(n) = t^n \qquad f(n) = u^n \qquad f(n) = nu^n.$$

Furthermore, every linear combination

$$f(n) = a \cdot s^n + b \cdot t^n + c \cdot u^n + d \cdot nu^n \tag{21.1}$$

is also a solution.

All that remains is to select a solution consistent with the boundary conditions by choosing the constants appropriately. Each boundary condition implies a linear equation involving these constants. So we can determine the constants by solving a system of linear equations. For example, suppose our boundary conditions were

$f(0) = 0$, $f(1) = 1$, $f(2) = 4$, and $f(3) = 9$. Then we would obtain four equations in four unknowns:

$$\begin{array}{lll} f(0) = 0 & \text{implies} & a \cdot s^0 + b \cdot t^0 + c \cdot u^0 + d \cdot 0u^0 = 0 \\ f(1) = 1 & \text{implies} & a \cdot s^1 + b \cdot t^1 + c \cdot u^1 + d \cdot 1u^1 = 1 \\ f(2) = 4 & \text{implies} & a \cdot s^2 + b \cdot t^2 + c \cdot u^2 + d \cdot 2u^2 = 4 \\ f(3) = 9 & \text{implies} & a \cdot s^3 + b \cdot t^3 + c \cdot u^3 + d \cdot 3u^3 = 9 \end{array}$$

This looks nasty, but remember that $s$, $t$, and $u$ are just constants. Solving this system gives values for $a$, $b$, $c$, and $d$ that define a solution to the recurrence consistent with the boundary conditions.

### 21.3.3 Solving General Linear Recurrences

We can now solve all linear homogeneous recurrences, which have the form

$$f(n) = a_1 f(n-1) + a_2 f(n-2) + \cdots + a_d f(n-d).$$

Many recurrences that arise in practice do not quite fit this mold. For example, the Towers of Hanoi problem led to this recurrence:

$$\begin{aligned} f(1) &= 1 \\ f(n) &= 2f(n-1) + 1 \qquad\qquad (\text{for } n \geq 2). \end{aligned}$$

The problem is the extra $+1$; that is not allowed in a homogeneous linear recurrence. In general, adding an extra function $g(n)$ to the right side of a linear recurrence gives an *inhomogeneous linear recurrence*:

$$f(n) = a_1 f(n-1) + a_2 f(n-2) + \cdots + a_d f(n-d) + g(n).$$

Solving inhomogeneous linear recurrences is neither very different nor very difficult. We can divide the whole job into five steps:

1. Replace $g(n)$ by 0, leaving a homogeneous recurrence. As before, find roots of the characteristic equation.

2. Write down the solution to the homogeneous recurrence, but do not yet use the boundary conditions to determine coefficients. This is called the *homogeneous solution*.

3. Now restore $g(n)$ and find a single solution to the recurrence, ignoring boundary conditions. This is called a *particular solution*. We'll explain how to find a particular solution shortly.

4. Add the homogeneous and particular solutions together to obtain the *general solution*.

5. Now use the boundary conditions to determine constants by the usual method of generating and solving a system of linear equations.

As an example, let's consider a variation of the Towers of Hanoi problem. Suppose that moving a disk takes time proportional to its size. Specifically, moving the smallest disk takes 1 second, the next-smallest takes 2 seconds, and moving the $n$th disk then requires $n$ seconds instead of 1. So, in this variation, the time to complete the job is given by a recurrence with a $+n$ term instead of a $+1$:

$$\begin{aligned} f(1) &= 1 \\ f(n) &= 2f(n-1) + n \qquad \text{for } n \geq 2. \end{aligned}$$

Clearly, this will take longer, but how much longer? Let's solve the recurrence with the method described above.

In Steps 1 and 2, dropping the $+n$ leaves the homogeneous recurrence $f(n) = 2f(n-1)$. The characteristic equation is $x = 2$. So the homogeneous solution is $f(n) = c2^n$.

In Step 3, we must find a solution to the full recurrence $f(n) = 2f(n-1) + n$, without regard to the boundary condition. Let's guess that there is a solution of the form $f(n) = an + b$ for some constants $a$ and $b$. Substituting this guess into the recurrence gives

$$\begin{aligned} an + b &= 2(a(n-1) + b) + n \\ 0 &= (a+1)n + (b - 2a). \end{aligned}$$

The second equation is a simplification of the first. The second equation holds for all $n$ if both $a + 1 = 0$ (which implies $a = -1$) and $b - 2a = 0$ (which implies that $b = -2$). So $f(n) = an + b = -n - 2$ is a particular solution.

In the Step 4, we add the homogeneous and particular solutions to obtain the general solution

$$f(n) = c2^n - n - 2.$$

Finally, in step 5, we use the boundary condition, $f(1) = 1$, determine the value of the constant $c$:

$$\begin{aligned} f(1) = 1 \quad &\text{IMPLIES} \quad c2^1 - 1 - 2 = 1 \\ &\text{IMPLIES} \quad c = 2. \end{aligned}$$

Therefore, the function $f(n) = 2 \cdot 2^n - n - 2$ solves this variant of the Towers of Hanoi recurrence. For comparison, the solution to the original Towers of Hanoi problem was $2^n - 1$. So if moving disks takes time proportional to their size, then the monks will need about twice as much time to solve the whole puzzle.

### 21.3.4 How to Guess a Particular Solution

Finding a particular solution can be the hardest part of solving inhomogeneous recurrences. This involves guessing, and you might guess wrong.[1] However, some rules of thumb make this job fairly easy most of the time.

- Generally, look for a particular solution with the same form as the inhomogeneous term $g(n)$.
- If $g(n)$ is a constant, then guess a particular solution $f(n) = c$. If this doesn't work, try polynomials of progressively higher degree: $f(n) = bn + c$, then $f(n) = an^2 + bn + c$, etc.
- More generally, if $g(n)$ is a polynomial, try a polynomial of the same degree, then a polynomial of degree one higher, then two higher, etc. For example, if $g(n) = 6n + 5$, then try $f(n) = bn + c$ and then $f(n) = an^2 + bn + c$.
- If $g(n)$ is an exponential, such as $3^n$, then first guess that $f(n) = c3^n$. Failing that, try $f(n) = bn3^n + c3^n$ and then $an^2 3^n + bn3^n + c3^n$, etc.

The entire process is summarized on the following page.

## 21.4 Divide-and-Conquer Recurrences

We now have a recipe for solving general linear recurrences. But the Merge Sort recurrence, which we encountered earlier, is not linear:

$$\begin{aligned} T(1) &= 0 \\ T(n) &= 2T(n/2) + n - 1 \qquad (\text{for } n \geq 2). \end{aligned}$$

In particular, $T(n)$ is not a linear combination of a fixed number of immediately preceding terms; rather, $T(n)$ is a function of $T(n/2)$, a term halfway back in the sequence.

[1] Chapter 16 explains how to solve linear recurrences with generating functions —it's a little more complicated, but it does not require guessing.

## Short Guide to Solving Linear Recurrences

A linear recurrence is an equation

$$f(n) = \underbrace{a_1 f(n-1) + a_2 f(n-2) + \cdots + a_d f(n-d)}_{\text{homogeneous part}} \quad \underbrace{+ g(n)}_{\text{inhomogeneous part}}$$

together with boundary conditions such as $f(0) = b_0$, $f(1) = b_1$, etc. Linear recurrences are solved as follows:

1. Find the roots of the characteristic equation
$$x^n = a_1 x^{n-1} + a_2 x^{n-2} + \cdots + a_{k-1} x + a_k.$$

2. Write down the homogeneous solution. Each root generates one term and the homogeneous solution is their sum. A nonrepeated root $r$ generates the term $c r^n$, where $c$ is a constant to be determined later. A root $r$ with multiplicity $k$ generates the terms
$$d_1 r^n \qquad d_2 n r^n \qquad d_3 n^2 r^n \qquad \ldots \qquad d_k n^{k-1} r^n$$
where $d_1, \ldots d_k$ are constants to be determined later.

3. Find a particular solution. This is a solution to the full recurrence that need not be consistent with the boundary conditions. Use guess-and-verify. If $g(n)$ is a constant or a polynomial, try a polynomial of the same degree, then of one higher degree, then two higher. For example, if $g(n) = n$, then try $f(n) = bn + c$ and then $an^2 + bn + c$. If $g(n)$ is an exponential, such as $3^n$, then first guess $f(n) = c3^n$. Failing that, try $f(n) = (bn + c)3^n$ and then $(an^2 + bn + c)3^n$, etc.

4. Form the general solution, which is the sum of the homogeneous solution and the particular solution. Here is a typical general solution:
$$f(n) = \underbrace{c2^n + d(-1)^n}_{\text{homogeneous solution}} + \underbrace{3n + 1}_{\text{inhomogeneous solution}}.$$

5. Substitute the boundary conditions into the general solution. Each boundary condition gives a linear equation in the unknown constants. For example, substituting $f(1) = 2$ into the general solution above gives
$$\begin{aligned} 2 &= c \cdot 2^1 + d \cdot (-1)^1 + 3 \cdot 1 + 1 \\ \text{IMPLIES} \quad -2 &= 2c - d. \end{aligned}$$
Determine the values of these constants by solving the resulting system of linear equations.

Merge Sort is an example of a divide-and-conquer algorithm: it divides the input, "conquers" the pieces, and combines the results. Analysis of such algorithms commonly leads to *divide-and-conquer* recurrences, which have this form:

$$T(n) = \sum_{i=1}^{k} a_i T(b_i n) + g(n)$$

Here $a_1, \ldots a_k$ are positive constants, $b_1, \ldots, b_k$ are constants between 0 and 1, and $g(n)$ is a nonnegative function. For example, setting $a_1 = 2$, $b_1 = 1/2$, and $g(n) = n - 1$ gives the Merge Sort recurrence.

### 21.4.1 The Akra-Bazzi Formula

The solution to virtually all divide and conquer solutions is given by the amazing *Akra-Bazzi formula*. Quite simply, the asymptotic solution to the general divide-and-conquer recurrence

$$T(n) = \sum_{i=1}^{k} a_i T(b_i n) + g(n)$$

is

$$T(n) = \Theta\left(n^p \left(1 + \int_1^n \frac{g(u)}{u^{p+1}}\, du\right)\right) \tag{21.2}$$

where $p$ satisfies

$$\sum_{i=1}^{k} a_i b_i^p = 1. \tag{21.3}$$

A rarely-troublesome requirement is that the function $g(n)$ must not grow or oscillate too quickly. Specifically, $|g'(n)|$ must be bounded by some polynomial. So, for example, the Akra-Bazzi formula is valid when $g(n) = x^2 \log n$, but not when $g(n) = 2^n$.

Let's solve the Merge Sort recurrence again, using the Akra-Bazzi formula instead of plug-and-chug. First, we find the value $p$ that satisfies

$$2 \cdot (1/2)^p = 1.$$

Looks like $p = 1$ does the job. Then we compute the integral:

$$\begin{aligned} T(n) &= \Theta\left(n\left(1 + \int_1^n \frac{u-1}{u^2}\,du\right)\right) \\ &= \Theta\left(n\left(1 + \left[\log u + \frac{1}{u}\right]_1^n\right)\right) \\ &= \Theta\left(n\left(\log n + \frac{1}{n}\right)\right) \\ &= \Theta(n \log n). \end{aligned}$$

The first step is integration and the second is simplification. We can drop the $1/n$ term in the last step, because the $\log n$ term dominates. We're done!

Let's try a scary-looking recurrence:

$$T(n) = 2T(n/2) + (8/9)T(3n/4) + n^2.$$

Here, $a_1 = 2$, $b_1 = 1/2$, $a_2 = 8/9$, and $b_2 = 3/4$. So we find the value $p$ that satisfies

$$2 \cdot (1/2)^p + (8/9)(3/4)^p = 1.$$

Equations of this form don't always have closed-form solutions, so you may need to approximate $p$ numerically sometimes. But in this case the solution is simple: $p = 2$. Then we integrate:

$$\begin{aligned} T(n) &= \Theta\left(n^2\left(1 + \int_1^n \frac{u^2}{u^3}\,du\right)\right) \\ &= \Theta\left(n^2(1 + \log n)\right) \\ &= \Theta\left(n^2 \log n\right). \end{aligned}$$

That was easy!

### 21.4.2 Two Technical Issues

Until now, we've swept a couple issues related to divide-and-conquer recurrences under the rug. Let's address those issues now.

First, the Akra-Bazzi formula makes no use of boundary conditions. To see why, let's go back to Merge Sort. During the plug-and-chug analysis, we found that

$$T_n = nT_1 + n \log n - n + 1.$$

This expresses the $n$th term as a function of the first term, whose value is specified in a boundary condition. But notice that $T_n = \Theta(n \log n)$ for *every* value of $T_1$. The boundary condition doesn't matter!

This is the typical situation: *the asymptotic solution to a divide-and-conquer recurrence is independent of the boundary conditions*. Intuitively, if the bottom-level operation in a recursive algorithm takes, say, twice as long, then the overall running time will at most double. This matters in practice, but the factor of 2 is concealed by asymptotic notation. There are corner-case exceptions. For example, the solution to $T(n) = 2T(n/2)$ is either $\Theta(n)$ or zero, depending on whether $T(1)$ is zero. These cases are of little practical interest, so we won't consider them further.

There is a second nagging issue with divide-and-conquer recurrences that does not arise with linear recurrences. Specifically, dividing a problem of size $n$ may create subproblems of non-integer size. For example, the Merge Sort recurrence contains the term $T(n/2)$. So what if $n$ is 15? How long does it take to sort seven-and-a-half items? Previously, we dodged this issue by analyzing Merge Sort only when the size of the input was a power of 2. But then we don't know what happens for an input of size, say, 100.

Of course, a practical implementation of Merge Sort would split the input *approximately* in half, sort the halves recursively, and merge the results. For example, a list of 15 numbers would be split into lists of 7 and 8. More generally, a list of $n$ numbers would be split into approximate halves of size $\lceil n/2 \rceil$ and $\lfloor n/2 \rfloor$. So the maximum number of comparisons is actually given by this recurrence:

$$\begin{aligned} T(1) &= 0 \\ T(n) &= T(\lceil n/2 \rceil) + T(\lfloor n/2 \rfloor) + n - 1 \qquad (\text{for } n \geq 2). \end{aligned}$$

This may be rigorously correct, but the ceiling and floor operations make the recurrence hard to solve exactly.

Fortunately, *the asymptotic solution to a divide and conquer recurrence is unaffected by floors and ceilings*. More precisely, the solution is not changed by replacing a term $T(b_i n)$ with either $T(ceil b_i n)$ or $T(\lfloor b_i n \rfloor)$. So leaving floors and ceilings out of divide-and-conquer recurrences makes sense in many contexts; those are complications that make no difference.

### 21.4.3 The Akra-Bazzi Theorem

The Akra-Bazzi formula together with our assertions about boundary conditions and integrality all follow from the *Akra-Bazzi Theorem*, which is stated below.

**Theorem 21.4.1** (Akra-Bazzi)**.** *Suppose that the function $T : \mathbb{R} \to \mathbb{R}$ satisfies the*

*recurrence*

$$T(x)\begin{cases}\text{is nonnegative and bounded} & \text{for } 0 \le x \le x_0,\\ = \sum_{i=1}^{k} a_i T(b_i x + h_i(x)) + g(x) & \text{for } x > x_0.\end{cases}$$

*where:*

1. $a_1, \ldots, a_k$ *are positive constants.*
2. $b_1, \ldots, b_k$ *are constants between 0 and 1.*
3. $x_0$ *is large enough so that* $T$ *is well-defined.*
4. $g(x)$ *is a nonnegative function such that* $|g'(x)|$ *is bounded by a polynomial.*
5. $|h_i(x)| = O(x/\log^2 x)$.

*Then*

$$T(x) = \Theta\left(x^p\left(1 + \int_1^x \frac{g(u)}{u^{p+1}}\,du\right)\right)$$

*where* $p$ *satisfies*

$$\sum_{i=1}^{k} a_i b_i^p = 1.$$

The Akra-Bazzi theorem can be proved using a complicated induction argument, though we won't do that here. But let's at least go over the statement of the theorem.

All the recurrences we've considered were defined over the integers, and that is the common case. But the Akra-Bazzi theorem applies more generally to functions defined over the real numbers.

The Akra-Bazzi formula is lifted directed from the theorem statement, except that the recurrence in the theorem includes extra functions, $h_i$. These functions extend the theorem to address floors, ceilings, and other small adjustments to the sizes of subproblems. The trick is illustrated by this combination of parameters

$$\begin{array}{lll} a_1 = 1 & b_1 = 1/2 & h_1(x) = \left\lceil \frac{x}{2} \right\rceil - \frac{x}{2} \\ a_2 = 1 & b_2 = 1/2 & h_2(x) = \left\lfloor \frac{x}{2} \right\rfloor - \frac{x}{2} \\ & g(x) = x - 1 & \end{array}$$

which corresponds the recurrence

$$\begin{aligned} T(x) &= 1 \cdot T\left(\frac{x}{2} + \left(\left\lceil \frac{x}{2} \right\rceil - \frac{x}{2}\right)\right) + \cdot T\left(\frac{x}{2} + \left(\left\lfloor \frac{x}{2} \right\rfloor - \frac{x}{2}\right)\right) + x - 1 \\ &= T\left(\left\lceil \frac{x}{2} \right\rceil\right) + T\left(\left\lfloor \frac{x}{2} \right\rfloor\right) + x - 1. \end{aligned}$$

This is the rigorously correct Merge Sort recurrence valid for all input sizes, complete with floor and ceiling operators. In this case, the functions $h_1(x)$ and $h_2(x)$ are both at most 1, which is easily $O(x/\log^2 x)$ as required by the theorem statement. These functions $h_i$ do not affect —or even appear in —the asymptotic solution to the recurrence. This justifies our earlier claim that applying floor and ceiling operators to the size of a subproblem does not alter the asymptotic solution to a divide-and-conquer recurrence.

### 21.4.4 The Master Theorem

There is a special case of the Akra-Bazzi formula known as the Master Theorem that handles some of the recurrences that commonly arise in computer science. It is called the *Master* Theorem because it was proved long before Akra and Bazzi arrived on the scene and, for many years, it was the final word on solving divide-and-conquer recurrences. We include the Master Theorem here because it is still widely referenced in algorithms courses and you can use it without having to know anything about integration.

**Theorem 21.4.2** (Master Theorem)**.** *Let $T$ be a recurrence of the form*

$$T(n) = aT\left(\frac{n}{b}\right) + g(n).$$

**Case 1:** *If $g(n) = O\left(n^{\log_b(a)-\epsilon}\right)$ for some constant $\epsilon > 0$, then*

$$T(n) = \Theta\left(n^{\log_b(a)}\right).$$

**Case 2:** *If $g(n) = \Theta\left(n^{\log_b(a)} \log^k(n)\right)$ for some constant $k \geq 0$, then*

$$T(n) = \Theta\left(n^{\log_b(a)} \log^{k+1}(n)\right).$$

**Case 3:** *If $g(n) = \Omega\left(n^{\log_b(a)+\epsilon}\right)$ for some constant $\epsilon > 0$ and $ag(n/b) < cg(n)$ for some constant $c < 1$ and sufficiently large $n$, then*

$$T(n) = \Theta(g(n)).$$

The Master Theorem can be proved by induction on $n$ or, more easily, as a corollary of Theorem 21.4.1. We will not include the details here.

## 21.5 A Feel for Recurrences

We've guessed and verified, plugged and chugged, found roots, computed integrals, and solved linear systems and exponential equations. Now let's step back and look for some rules of thumb. What kinds of recurrences have what sorts of solutions?

Here are some recurrences we solved earlier:

| | Recurrence | Solution |
|---|---|---|
| Towers of Hanoi | $T_n = 2T_{n-1} + 1$ | $T_n \sim 2^n$ |
| Merge Sort | $T_n = 2T_{n/2} + n - 1$ | $T_n \sim n \log n$ |
| Hanoi variation | $T_n = 2T_{n-1} + n$ | $T_n \sim 2 \cdot 2^n$ |
| Fibonacci | $T_n = T_{n-1} + T_{n-2}$ | $T_n \sim (1.618\ldots)^{n+1}/\sqrt{5}$ |

Notice that the recurrence equations for Towers of Hanoi and Merge Sort are somewhat similar, but the solutions are radically different. Merge Sorting $n = 64$ items takes a few hundred comparisons, while moving $n = 64$ disks takes more than $10^{19}$ steps!

Each recurrence has one strength and one weakness. In the Towers of Hanoi, we broke a problem of size $n$ into two subproblem of size $n - 1$ (which is large), but needed only 1 additional step (which is small). In Merge Sort, we divided the problem of size $n$ into two subproblems of size $n/2$ (which is small), but needed $(n - 1)$ additional steps (which is large). Yet, Merge Sort is faster by a mile!

This suggests that *generating smaller subproblems is far more important to algorithmic speed than reducing the additional steps per recursive call*. For example, shifting to the variation of Towers of Hanoi increased the last term from $+1$ to $+n$, but the solution only doubled. And one of the two subproblems in the Fibonacci recurrence is just *slightly* smaller than in Towers of Hanoi (size $n - 2$ instead of $n-1$). Yet the solution is exponentially smaller! More generally, linear recurrences (which have big subproblems) typically have exponential solutions, while divide-and-conquer recurrences (which have small subproblems) usually have solutions bounded above by a polynomial.

All the examples listed above break a problem of size $n$ into two smaller problems. How does the number of subproblems affect the solution? For example, suppose we increased the number of subproblems in Towers of Hanoi from 2 to 3, giving this recurrence:

$$T_n = 3T_{n-1} + 1$$

This increases the root of the characteristic equation from 2 to 3, which raises the solution exponentially, from $\Theta(2^n)$ to $\Theta(3^n)$.

Divide-and-conquer recurrences are also sensitive to the number of subproblems. For example, for this generalization of the Merge Sort recurrence:

$$\begin{aligned} T_1 &= 0 \\ T_n &= aT_{n/2} + n - 1. \end{aligned}$$

the Akra-Bazzi formula gives:

$$T_n = \begin{cases} \Theta(n) & \text{for } a < 2 \\ \Theta(n \log n) & \text{for } a = 2 \\ \Theta(n^{\log a}) & \text{for } a > 2. \end{cases}$$

So the solution takes on three completely different forms as $a$ goes from 1.99 to 2.01!

How do boundary conditions affect the solution to a recurrence? We've seen that they are almost irrelevant for divide-and-conquer recurrences. For linear recurrences, the solution is usually dominated by an exponential whose base is determined by the number and size of subproblems. Boundary conditions matter greatly only when they give the dominant term a zero coefficient, which changes the asymptotic solution.

So now we have a rule of thumb! The performance of a recursive procedure is usually dictated by the size and number of subproblems, rather than the amount of work per recursive call or time spent at the base of the recursion. In particular, if subproblems are smaller than the original by an additive factor, the solution is most often exponential. But if the subproblems are only a fraction the size of the original, then the solution is typically bounded by a polynomial.

# 11 Simple Graphs

*Simple graphs* model relationships that are *symmetric*, meaning that the relationship is mutual. Examples of such mutual relationships are being married, speaking the same language, not speaking the same language, occurring during overlapping time intervals, or being connected by a conducting wire. They come up in all sorts of applications, including scheduling, constraint satisfaction, computer graphics, and communications, but we'll start with an application designed to get your attention: we are going to make a professional inquiry into sexual behavior. Namely, we'll look at some data about who, on average, has more opposite-gender partners, men or women.

Sexual demographics have been the subject of many studies. In one of the largest studies, researchers from the University of Chicago interviewed a random sample of 2500 people over several years to try to get an answer to this question. Their study, published in 1994, and entitled *The Social Organization of Sexuality* found that on average men have 74% more opposite-gender partners than women.

Other studies have found that the disparity is even larger. In particular, ABC News claimed that the average man has 20 partners over his lifetime, and the average woman has 6, for a percentage disparity of 233%. The ABC News study, aired on Primetime Live in 2004, purported to be one of the most scientific ever done, with only a 2.5% margin of error. It was called "American Sex Survey: A peek between the sheets," —which raises some questions about the seriousness of their reporting.

Yet again, in August, 2007, the N.Y. Times reported on a study by the National Center for Health Statistics of the U.S. government showing that men had seven partners while women had four. Anyway, whose numbers do you think are more accurate, the University of Chicago, ABC News, or the National Center? —don't answer; this is a setup question like "When did you stop beating your wife?" Using a little graph theory, we'll explain why none of these findings can be anywhere near the truth.

## 11.1 Vertex Adjacency and Degrees

Simple graphs are defined as digraphs in which edges are *undirected* —they just connect two vertices without pointing in either direction between the vertices. So instead of a directed edge $\langle v \to w \rangle$ which starts at vertex $v$ and ends at vertex $w$, a

simple graph only has an undirected edge, $\langle v—w\rangle$, that connects $v$ and $w$.

**Definition 11.1.1.** A *simple graph*, $G$, consists of a nonempty set, $V(G)$, called the *vertices* of $G$, and a set $E(G)$ called the *edges* of $G$. An element of $V(G)$ is called a *vertex*. A vertex is also called a *node*; the words "vertex" and "node" are used interchangeably. An element of $E(G)$ an *undirected edge* or simply an "edge." An undirected edge has two vertices $u \neq v$ called its *endpoints*. Such an edge can be represented by the two element set $\{u, v\}$. The notation $\langle u—v\rangle$ denotes this edge.

Both $\langle u—v\rangle$ and $\langle v—u\rangle$ define the same undirected edge, namely the one whose endpoints are $u$ and $v$.

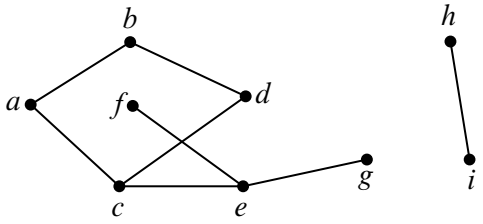



**Figure 11.1** An example of a graph with 9 nodes and 8 edges.

For example, let $H$ be the graph pictured in Figure 11.1. The vertices of $H$ correspond to the nine dots in Figure 11.1, that is,

$$V(H) = \{a, b, c, d, e, f, g, h, i\}.$$

The edges correspond to the eight lines, that is,

$$E(H) = \{\langle a—b\rangle, \langle a—c\rangle, \langle b—d\rangle, \langle c—d\rangle, \langle c—e\rangle, \langle e—f\rangle, \langle e—g\rangle, \langle h—i\rangle\}.$$

Mathematically, that's all there is to the graph $H$.

**Definition 11.1.2.** Two vertices in a simple graph are said to be *adjacent* iff they are the endpoints of the same edge, and an edge is said to be *incident* to each of its endpoints. The number of edges incident to a vertex $v$ is called the *degree* of the vertex and is denoted by $\deg(v)$. Equivalently, the degree of a vertex is the number of vertices adjacent to it.

For example, for the graph $H$ of Figure 11.1, vertex $a$ is adjacent to vertex $b$, and $b$ is adjacent to $d$. The edge $\langle a—c\rangle$ is incident to its endpoints $a$ and $c$. Vertex $h$ has degree 1, $d$ has degree 2, and $\deg(e) = 3$. It is possible for a vertex to have degree 0, in which case it is not adjacent to any other vertices. A simple graph, $G$, does not need to have any edges at all, namely $|E(G)|$ could be zero, which implies

that the degree of every vertex is also zero. But a simple graph must have at least one vertex, that is, $|V(G)|$ is required to be at least one.

An edge whose endpoints are the same is called a *self-loop*. Self-loops aren't allowed in simple graphs.[1] In a more general class of graphs called *multigraphs* there can be more than one edge with the same two endpoints, but this doesn't happen in simple graphs since every edge is uniquely determined by its two endpoints.

Sometimes graphs with no vertices, with self-loops, or with more than one edge between the same two vertices are convenient to have, but we don't need them, and sticking with simple graphs is simpler. :-)

*For the rest of this chapter we'll use "graphs" as an abbreviation for "simple graphs."*

A synonym for "vertices" is "*nodes*," and we'll use these words interchangeably. Simple graphs are sometimes called *networks*, edges are sometimes called *arcs*. We mention this as a "heads up" in case you look at other graph theory literature; we won't use these words.

## 11.2 Sexual Demographics in America

Let's model the question of heterosexual partners in graph theoretic terms. To do this, we'll let $G$ be the graph whose vertices, $V$, are all the people in America. Then we split $V$ into two separate subsets: $M$, which contains all the males, and $F$, which contains all the females.[2] We'll put an edge between a male and a female iff they have been sexual partners. This graph is pictured in Figure 11.2 with males on the left and females on the right.

Actually, this is a pretty hard graph to figure out, let alone draw. The graph is *enormous*: the US population is about 300 million, so $|V| \approx 300M$. Of these, approximately 50.8% are female and 49.2% are male, so $|M| \approx 147.6M$, and $|F| \approx 152.4M$. And we don't even have trustworthy estimates of how many edges there are, let alone exactly which couples are adjacent. But it turns out that we don't need to know any of this —we just need to figure out the relationship between the average number of partners per male and partners per female. To do this, we note that every edge has exactly one endpoint in $M$ vertex (remember, we're only considering male-female relationships); so the sum of the degrees of the $M$ vertices equals the number of edges. For the same reason, the sum of the degrees of the $F$

[1] You might try to represent a self-loop going between a vertex $v$ and itself as $\{v, v\}$, but this equals $\{v\}$, and it wouldn't be an edge which is defined to be a set of *two* vertices.

[2] For simplicity, we'll ignore the possibility of someone being *both* a man and a woman, or neither.

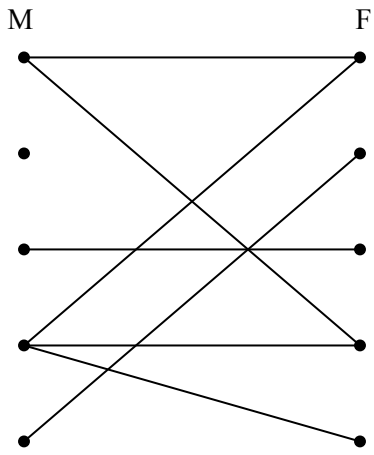



**Figure 11.2** The sex partners graph.

vertices equals the number of edges. So these sums are equal:

$$\sum_{x \in M} \deg(x) = \sum_{y \in F} \deg(y).$$

Now suppose we divide both sides of this equation by the product of the sizes of the two sets, $|M| \cdot |F|$:

$$\left(\frac{\sum_{x \in M} \deg(x)}{|M|}\right) \cdot \frac{1}{|F|} = \left(\frac{\sum_{y \in F} \deg(y)}{|F|}\right) \cdot \frac{1}{|M|}$$

The terms above in parentheses are the *average degree of an $M$ vertex* and the *average degree of a $F$* vertex. So we know:

$$\text{Avg. deg in } M = \frac{|F|}{|M|} \cdot \text{Avg. deg in } F \tag{11.1}$$

In other words, we've proved that the average number of female partners of males in the population compared to the average number of males per female is *determined solely by the relative number of males and females in the population.*

Now the Census Bureau reports that there are slightly more females than males in America; in particular $|F|/|M|$ is about 1.035. So we know that on average, males have 3.5% more opposite-gender partners than females, and this tells us nothing about any sex's promiscuity or selectivity. Rather, it just has to do with the relative number of males and females. Collectively, males and females have the same number of opposite gender partners, since it takes one of each set for every partnership,

but there are fewer males, so they have a higher ratio. This means that the University of Chicago, ABC, and the Federal government studies are way off. After a huge effort, they gave a totally wrong answer.

There's no definite explanation for why such surveys are consistently wrong. One hypothesis is that males exaggerate their number of partners —or maybe females downplay theirs —but these explanations are speculative. Interestingly, the principal author of the National Center for Health Statistics study reported that she knew the results had to be wrong, but that was the data collected, and her job was to report it.

The same underlying issue has led to serious misinterpretations of other survey data. For example, a couple of years ago, the Boston Globe ran a story on a survey of the study habits of students on Boston area campuses. Their survey showed that on average, minority students tended to study with non-minority students more than the other way around. They went on at great length to explain why this "remarkable phenomenon" might be true. But it's not remarkable at all —using our graph theory formulation, we can see that all it says is that there are fewer minority students than non-minority students, which is, of course what "minority" means.

### 11.2.1 Handshaking Lemma

The previous argument hinged on the connection between a sum of degrees and the number edges. There is a simple connection between these in any graph:

**Lemma 11.2.1.** *The sum of the degrees of the vertices in a graph equals twice the number of edges.*

*Proof.* Every edge contributes two to the sum of the degrees, one for each of its endpoints. ■

Lemma 11.2.1 is sometimes called the *Handshake Lemma*: if we total up the number of people each person at a party shakes hands with, the total will be twice the number of handshakes that occurred.

## 11.3 Some Common Graphs

Some graphs come up so frequently that they have names. A *complete graph* $K_n$ has $n$ vertices and an edge between every two vertices, for a total of $n(n-1)/2$ edges. For example, $K_5$ is shown in Figure 11.3.

The *empty graph* has no edges at all. For example, the empty graph with 5 nodes is shown in Figure 11.4.

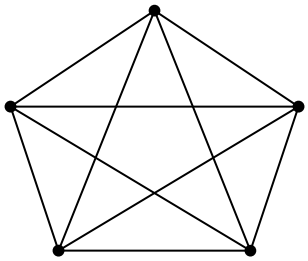


**Figure 11.3** $K_5$: the complete graph on 5 nodes.

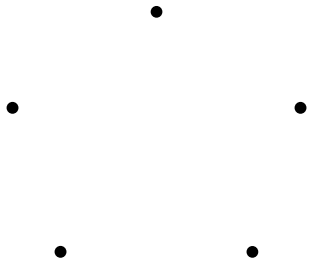


**Figure 11.4** An empty graph with 5 nodes.

An $n$-node graph containing $n-1$ edges in sequence is known as a *line graph* $L_n$. More formally, $L_n$ has

$$V(L_n) = \{v_1, v_2, \ldots, v_n\}$$

and

$$E(L_n) = \{\langle v_1 — v_2\rangle, \langle v_2 — v_3\rangle, \ldots, \langle v_{n-1} — v_n\rangle\}$$

For example, $L_5$ is pictured in Figure 11.5.

There is also a one-way infinite line graph $L_\infty$ which can be defined by letting the nonnegative integers $\mathbb{N}$ be the vertices with edges $\langle k — (k+1)\rangle$ for all $k \in \mathbb{N}$.

If we add the edge $\langle v_n — v_1\rangle$ to the line graph $L_n$, we get a graph called a *length-$n$ cycle* $C_n$. Figure 11.6 shows a picture of length-5 cycle.

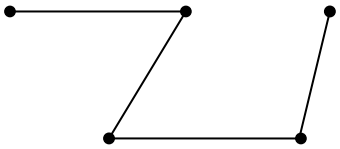


**Figure 11.5** $L_5$: a 5-node line graph.

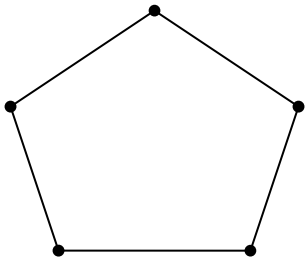


**Figure 11.6** $C_5$: a 5-node cycle graph.

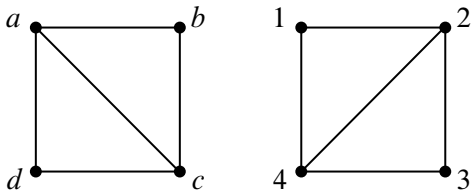



**Figure 11.7** Two Isomorphic graphs.

## 11.4 Isomorphism

Two graphs that look the same might actually be different in a formal sense. For example, the two graphs in Figure 11.7 are both 4-vertex, 5-edge graphs and you get graph (b) by a 90° clockwise rotation of graph (a).

Strictly speaking, these graphs are different mathematical objects, but this difference doesn't reflect the fact that the two graphs can be described by the same picture—except for the labels on the vertices. This idea of having the same picture "up to relabeling" can be captured neatly by adapting Definition 9.7.1 of isomorphism of digraphs to handle simple graphs. An isomorphism between two graphs is an edge-preserving bijection between their sets of vertices:

**Definition 11.4.1.** An isomorphism between graphs $G$ and $H$ is a bijection $f : V(G) \to V(H)$ such that

$$\langle u—v\rangle \in E(G) \quad \text{iff} \quad \langle f(u)—f(v)\rangle \in E(H)$$

for all $u, v \in V(G)$. Two graphs are isomorphic when there is an isomorphism between them.

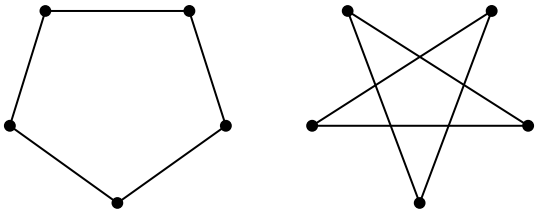


**Figure 11.8** Isomorphic $C_5$ graphs.

Here is an isomorphism, $f$, between the two graphs in Figure 11.7:

$$\begin{aligned} f(a) &::= 2 & f(b) &::= 3 \\ f(c) &::= 4 & f(d) &::= 1. \end{aligned}$$

You can check that there is an edge between two vertices in the graph on the left if and only if there is an edge between the two corresponding vertices in the graph on the right.

Two isomorphic graphs may be drawn very differently. For example, Figure 11.8 shows two different ways of drawing $C_5$:

Notice that if $f$ is an isomorphism between $G$ and $H$, then $f^{-1}$ is an isomorphism between $H$ and $G$. Isomorphism is also transitive because the composition of isomorphisms is an isomorphism. So isomorphism is in fact an equivalence relation.

Isomorphism preserves the connection properties of a graph, abstracting out what the vertices are called, what they are made out of, or where they appear in a drawing of the graph. More precisely, a property of a graph is said to be *preserved under isomorphism* if whenever $G$ has that property, every graph isomorphic to $G$ also has that property. For example, since an isomorphism is a bijection between sets of vertices, isomorphic graphs must have the same number of vertices. What's more, if $f$ is a graph isomorphism that maps a vertex, $v$, of one graph to the vertex, $f(v)$, of an isomorphic graph, then by definition of isomorphism, every vertex adjacent to $v$ in the first graph will be mapped by $f$ to a vertex adjacent to $f(v)$ in the isomorphic graph. That is, $v$ and $f(v)$ will have the same degree. So if one graph has a vertex of degree 4 and another does not, then they can't be isomorphic. In fact, they can't be isomorphic if the number of degree 4 vertices in each of the graphs is not the same.

Looking for preserved properties can make it easy to determine that two graphs are not isomorphic, or to guide the search for an isomorphism when there is one. It's generally easy in practice to decide whether two graphs are isomorphic. However, no one has yet found a procedure for determining whether two graphs are

isomorphic that is *guaranteed* to run in polynomial time on all pairs of graphs.[3]

Having such a procedure would be useful. For example, it would make it easy to search for a particular molecule in a database given the molecular bonds. On the other hand, knowing there is no such efficient procedure would also be valuable: secure protocols for encryption and remote authentication can be built on the hypothesis that graph isomorphism is computationally exhausting.

The definitions of bijection and isomorphism apply infinite graphs as well as finite graphs, as do most of the results in the rest of this chapter. But graph theory focuses mostly on finite graphs, and we will too. So

*in the rest of this chapter we'll assume graphs are finite.*

We've actually been taking isomorphism for granted ever since we wrote "$K_n$ has $n$ vertices..." at the beginning of section 11.3.

*Graph theory is all about properties preserved by isomorphism.*

## 11.5 Bipartite Graphs & Matchings

There were two kinds of vertices in the "Sex in America" graph —males and females, and edges only went between the two kinds. Graphs like this come up so frequently that they have earned a special name —they are called *bipartite graphs*.

**Definition 11.5.1.** A *bipartite graph* is a graph whose vertices can be partitioned[4] into two sets, $L(G)$ and $R(G)$, such that every edge has one endpoint in $L(G)$ and the other endpoint in $R(G)$.

So every bipartite graph looks something like the graph in Figure 11.2.

### 11.5.1 The Bipartite Matching Problem

The bipartite matching problem is related to the sex-in-America problem that we just studied; only now the goal is to get everyone happily married. As you might imagine, this is not possible for a variety of reasons, not the least of which is the fact that there are more women in America than men. So, it is simply not possible to marry every woman to a man so that every man is married at most once.

But what about getting a mate for every man so that every woman is married at most once? Is it possible to do this so that each man is paired with a woman that

[3] A procedure runs in *polynomial time* when it needs an amount of time of at most $p(n)$, where $n$ is the total number of vertices and $p()$ is a fixed polynomial.

[4] Partitioning a set means cutting it up into *nonempty* pieces. In this case, it means that $L(G)$ and $R(G)$ are nonempty, $L(G) \cup R(G) = V(G)$, and $L(G) \cap R(G) = \emptyset$.

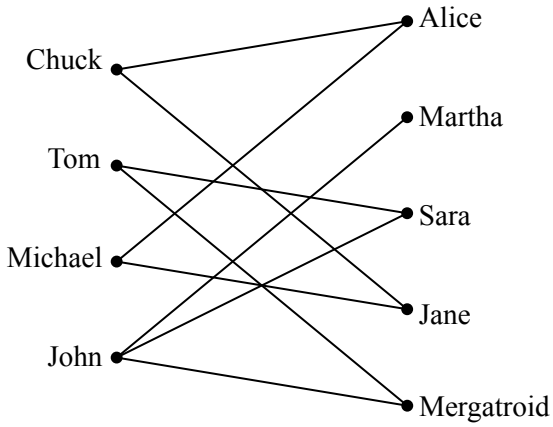



**Figure 11.9** A graph where an edge between a man and woman denotes that the man likes the woman.

he likes? The answer, of course, depends on the bipartite graph that represents who likes who, but the good news is that it is possible to find natural properties of the who-likes-who graph that completely determine the answer to this question.

In general, suppose that we have a set of men and an equal-sized or larger set of women, and there is a graph with an edge between a man and a woman if the man likes the woman. In this scenario, the "likes" relationship need not be symmetric, since for the time being, we will only worry about finding a mate for each man that he likes.[5] (Later, we will consider the "likes" relationship from the female perspective as well.) For example, we might obtain the graph in Figure 11.9.

A *matching* is defined to be an assignment of a woman to each man so that different men are assigned to different women, and a man is always assigned a woman that he likes. For example, one possible matching for the men is shown in Figure 11.10.

### The Matching Condition

A famous result known as Hall's Matching Theorem gives necessary and sufficient conditions for the existence of a matching in a bipartite graph. It turns out to be a remarkably useful mathematical tool.

We'll state and prove Hall's Theorem using man-likes-woman terminology. Define *the set of women liked by a given set of men* to consist of all women liked by

[5] By the way, we do not mean to imply that marriage should or should not be of a heterosexual nature. Nor do we mean to imply that men should get their choice instead of women. It's just that with bipartite graphs, the edges only connected male nodes to female nodes and there are fewer men in America. So please don't take offense.

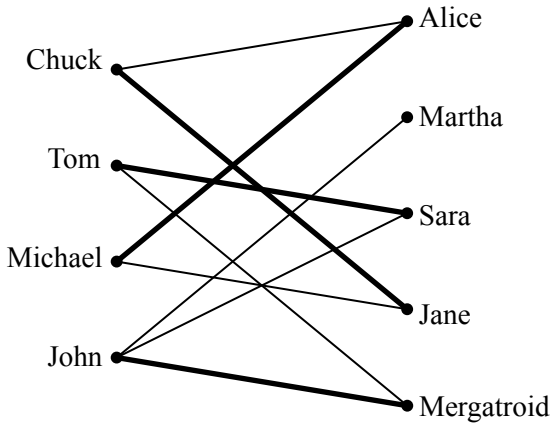



**Figure 11.10** One possible matching for the men is shown with bold edges. For example, John is matched with Mergatroid.

at least one of those men. For example, the set of women liked by Tom and John in Figure 11.9 consists of Martha, Sara, and Mergatroid. For us to have any chance at all of matching up the men, the following *matching condition* must hold:

*The Matching Condition*: every subset of men likes at least as large a set of women.

For example, we can not find a matching if some set of 4 men like only 3 women. Hall's Theorem says that this necessary condition is actually sufficient; if the matching condition holds, then a matching exists.

**Theorem 11.5.2.** *A matching for a set $M$ of men with a set $W$ of women can be found if and only if the matching condition holds.*

*Proof.* First, let's suppose that a matching exists and show that the matching condition holds. For any subset of men, each man likes at least the woman he is matched with and a woman is matched with at most one man. Therefore, every subset of men likes at least as large a set of women. Thus, the matching condition holds.

Next, let's suppose that the matching condition holds and show that a matching exists. We use strong induction on $|M|$, the number of men, on the predicate:

$$P(m) ::= \text{ if the matching condition holds for a set, } M, \text{ of } m \text{ men, then there is a matching for } M.$$

**Base case** ($|M| = 1$): If $|M| = 1$, then the matching condition implies that the lone man likes at least one woman, and so a matching exists.

**Inductive Step**: Suppose that $|M| = m + 1 \geq 2$. To find a matching for $M$, there are two cases.

**Case 1:** Every nonempty subset of at most $m$ men likes a *strictly larger* set of women. In this case, we have some latitude: we pair an arbitrary man with a woman he likes and send them both away. This leaves $m$ men and one fewer women, and the matching condition will still hold. So the induction hypothesis $P(m)$ implies we can match the remaining $m$ men.

**Case 2:** Some nonempty subset, $X$, of at most $m$ men likes an *equal-size* set, $Y$, of women. The matching condition must hold within $X$, so the strong induction hypothesis implies we can match the men in $X$ with the women in $Y$. This leaves the problem of matching the set $M - X$ of men to the set $W - Y$ of women.

But the problem of matching $M - X$ against $W - Y$ also satisfies the bottleneck condition, because any bottleneck for $M - X$ would imply a bottleneck within the original set of men, $M$. Namely, if a subset $M_0 \subseteq M - X$ liked only a strictly smaller subset of women $W_0 \subseteq W - Y$, then the set $M_0 \cup X$ of men would like only women in the strictly smaller set $W_0 \cup Y$. So again the strong induction hypothesis implies we can match the men in $M - X$ with the women in $W - Y$, which completes a matching for $M$.

So in both cases, there is a matching for the men, which completes the proof of the Inductive step. The theorem follows by induction. ■

The proof of Theorem 11.5.2 gives an algorithm for finding a matching in a bipartite graph, albeit not a very efficient one. However, efficient algorithms for finding a matching in a bipartite graph do exist. Thus, if a problem can be reduced to finding a matching, the problem is essentially solved from a computational perspective.

### A Formal Statement

Let's restate Theorem 11.5.2 in abstract terms so that you'll not always be condemned to saying, "Now this group of men likes at least as many women..."

**Definition 11.5.3.** A *matching* in a graph $G$ is a set $M$ of edges of $G$ such that no vertex is an endpoint of more than one edge in $M$. A matching is said to *cover* a set, $S$, of vertices iff each vertex in $S$ is an endpoint of an edge of the matching. A matching is said to be *perfect* if it covers $V(G)$. In any graph, the set $N(S)$ of *neighbors* of some set $S$ of vertices is the image of $S$ under the edge-relation, that is,

$$N(S) ::= \{ r \mid \langle s\text{—}r \rangle \in E(G) \text{ for some } s \in S \}.$$

$S$ is called a *bottleneck* if

$$|S| > |N(S)|.$$

**Theorem 11.5.4** (Hall's Theorem)**.** *Let $G$ be a bipartite graph. There is a matching in $G$ that covers $L(G)$ iff no subset of $L(G)$ is a bottleneck.*

### An Easy Matching Condition

The bipartite matching condition requires that *every* subset of men has a certain property. In general, verifying that every subset has some property, even if it's easy to check any particular subset for the property, quickly becomes overwhelming because the number of subsets of even relatively small sets is enormous—over a billion subsets for a set of size 30. However, there is a simple property of vertex degrees in a bipartite graph that guarantees the existence of a matching. Namely, call a bipartite graph *degree-constrained* if vertex degrees on the left are at least as large as those on the right. More precisely,

**Definition 11.5.5.** A bipartite graph $G$ is *degree-constrained* when $\deg(l) \geq \deg(r)$ for every $l \in L(G)$ and $r \in R(G)$.

For example, the graph in Figure 11.9 is degree-constrained since every node on the left is adjacent to at least two nodes on the right while every node on the right is adjacent to at most two nodes on the left.

**Theorem 11.5.6.** *If $G$ is a degree-constrained bipartite graph, then there is a matching that covers $L(G)$.*

*Proof.* We will show that $G$ satisfies Hall's condition, namely, if $S$ is an arbitrary subset of $L(G)$, then

$$|N(S)| \geq |S|. \tag{11.2}$$

Since $G$ is degree-constrained, there is a $d > 0$ such that $\deg(l) \geq d \geq \deg(r)$ for every $l \in L$ and $r \in R$. Since every edge with an endpoint in $S$ has its other endpoint in $N(S)$ by definition, and every node in $N(S)$ is incident to at most $d$ edges, we know that

$$d|N(S)| \geq \#\text{edges with an endpoint in } S.$$

Also, since every node in $S$ is the endpoint of at least $d$ edges,

$$\#\text{edges incident to a vertex in } S \geq d|S|.$$

It follows that $d|N(S)| \geq d|S|$. Cancelling $d$ completes the derivation of equation (11.2). ■

Regular graphs are a large class of degree-constrained graphs that often arise in practice. Hence, we can use Theorem 11.5.6 to prove that every regular bipartite graph has a perfect matching. This turns out to be a surprisingly useful result in computer science.

**Definition 11.5.7.** A graph is said to be *regular* if every node has the same degree.

**Theorem 11.5.8.** *Every regular bipartite graph has a perfect matching.*

*Proof.* Let $G$ be a regular bipartite graph. Since regular graphs are degree-constrained, we know by Theorem 11.5.6 that there must be a matching in $G$ that covers $L(G)$. Such a matching is only possible when $|L(G)| \leq |R(G)|$. But $G$ is also degree-constrained if the roles of $L(G)$ and $R(G)$ are switched, which implies that $|R(G)| \leq |L(G)|$ also. That is, $L(G)$ and $R(G)$ are the same size, and any matching covering $L(G)$ will also cover $R(G)$. So every node in $G$ is an endpoint of an edge in the matching, and thus $G$ has a perfect matching. ■

## 11.6 The Stable Marriage Problem

We next consider a version of the bipartite matching problem where there are an equal number of men and women, and where each person has preferences about who they would like to marry. In fact, we assume that each man has a complete list of all the women ranked according to his preferences, with no ties. Likewise, each woman has a ranked list of all of the men.

The preferences don't have to be symmetric. That is, Jennifer might like Brad best, but Brad doesn't necessarily like Jennifer best. The goal is to marry everyone: every man must marry exactly one woman and vice-versa—no polygamy. Moreover, we would like to find a matching between men and women that is *stable* in the sense that there is no pair of people that prefer each other to their spouses.

For example, suppose *every* man likes Angelina best, and every woman likes Brad best, but Brad and Angelina are married to other people, say Jennifer and Billy Bob. Now *Brad and Angelina prefer each other to their spouses*, which puts their marriages at risk: pretty soon, they're likely to start spending late nights together working on problem sets!

This unfortunate situation is illustrated in Figure 11.11, where the digits "1" and "2" near a man shows which of the two women he ranks first and second, respectively, and similarly for the women.

More generally, in any matching, a man and woman who are not married to each other and who like each other better than their spouses, is called a *rogue couple*. In the situation shown in Figure 11.11, Brad and Angelina would be a rogue couple.

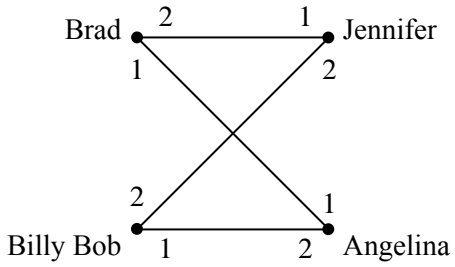



**Figure 11.11** Preferences for four people. Both men like Angelina best and both women like Brad best.

Having a rogue couple is not a good thing, since it threatens the stability of the marriages. On the other hand, if there are no rogue couples, then for any man and woman who are not married to each other, at least one likes their spouse better than the other, and so they won't be tempted to start an affair.

**Definition 11.6.1.** A *stable matching* is a matching with no rogue couples.

The question is, given everybody's preferences, how do you find a stable set of marriages? In the example consisting solely of the four people in Figure 11.11, we could let Brad and Angelina both have their first choices by marrying each other. Now neither Brad nor Angelina prefers anybody else to their spouse, so neither will be in a rogue couple. This leaves Jen not-so-happily married to Billy Bob, but neither Jen nor Billy Bob can entice somebody else to marry them, and so there is a stable matching.

Surprisingly, there always is a stable matching among a group of men and women. The surprise springs in part from considering the apparently similar "buddy" matching problem. That is, if people can be paired off as buddies, regardless of gender, then a stable matching *may not* be possible. For example, Figure 11.12 shows a situation with a love triangle and a fourth person who is everyone's last choice. In this figure Mergatroid's preferences aren't shown because they don't even matter. Let's see why there is no stable matching.

**Lemma 11.6.2.** *There is no stable buddy matching among the four people in Figure 11.12.*

*Proof.* We'll prove this by contradiction.

Assume, for the purposes of contradiction, that there is a stable matching. Then there are two members of the love triangle that are matched. Since preferences in the triangle are symmetric, we may assume in particular, that Robin and Alex are matched. Then the other pair must be Bobby-Joe matched with Mergatroid.

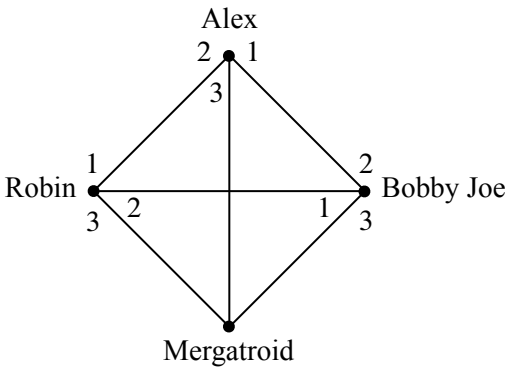



**Figure 11.12** Some preferences with no stable buddy matching.

But then there is a rogue couple: Alex likes Bobby-Joe best, and Bobby-Joe prefers Alex to his buddy Mergatroid. That is, Alex and Bobby-Joe are a rogue couple, contradicting the assumed stability of the matching. ■

So getting a stable *buddy* matching may not only be hard, it may be impossible. But when men are only allowed to marry women, and vice versa, then it turns out that a stable matching can always be found.[6]

### 11.6.1 The Mating Ritual

The procedure for finding a stable matching involves a *Mating Ritual* that takes place over several days. The following events happen each day:

**Morning**: Each woman stands on her balcony. Each man stands under the balcony of his favorite among the women on his list, and he serenades her. If a man has no women left on his list, he stays home and does his math homework.

**Afternoon**: Each woman who has one or more suitors serenading her, says to her favorite among them, "We might get engaged. Come back tomorrow." To the other suitors, she says, "No. I will never marry you! Take a hike!"

**Evening**: Any man who is told by a woman to take a hike, crosses that woman off his list.

**Termination condition**: When a day arrives in which every woman has at most one suitor, the ritual ends with each woman marrying her suitor, if she has one.

There are a number of facts about this Mating Ritual that we would like to prove:

- The Ritual eventually reaches the termination condition.

[6] Once again, we disclaim any political statement here —it's just the way that the math works out.

- Everybody ends up married.
- The resulting marriages are stable.

### 11.6.2 There is a Marriage Day

It's easy to see why the Mating Ritual has a terminal day when people finally get married. Every day on which the ritual hasn't terminated, at least one man crosses a woman off his list. (If the ritual hasn't terminated, there must be some woman serenaded by at least two men, and at least one of them will have to cross her off his list). If we start with $n$ men and $n$ women, then each of the $n$ men's lists initially has $n$ women on it, for a total of $n^2$ list entries. Since no women ever gets added to a list, the total number of entries on the lists decreases every day that the Ritual continues, and so the Ritual can continue for at most $n^2$ days.

### 11.6.3 They All Live Happily Ever After...

We still have to prove that the Mating Ritual leaves everyone in a stable marriage. To do this, we note one very useful fact about the Ritual: if a woman has a favorite suitor on some morning of the Ritual, then that favorite suitor will still be serenading her the next morning—because his list won't have changed. So she is sure to have today's favorite man among her suitors tomorrow. That means she will be able to choose a favorite suitor tomorrow who is at least as desirable to her as today's favorite. So day by day, her favorite suitor can stay the same or get better, never worse. This sounds like an invariant, and it is.

**Definition 11.6.3.** Let $P$ be the predicate: For every woman, $w$, and every man, $m$, if $w$ is crossed off $m$'s list, then $w$ has a suitor whom she prefers over $m$.

**Lemma 11.6.4.** *$P$ is an invariant for The Mating Ritual.*

*Proof.* By induction on the number of days.

**Base case**: In the beginning—that is, at the end of day 0—every woman is on every list. So no one has been crossed off, and $P$ is vacuously true.

**Inductive Step**: Assume $P$ is true at the end of day $d$ and let $w$ be a woman that has been crossed off a man $m$'s list by the end of day $d+1$.

**Case 1:** $w$ was crossed off $m$'s list on day $d+1$. Then, $w$ must have a suitor she prefers on day $d+1$.

**Case 2:** $w$ was crossed off $m$'s list prior to day $d+1$. Since $P$ is true at the end of day $d$, this means that $w$ has a suitor she prefers to $m$ on day $d$. She therefore has the same suitor or someone she prefers better at the end of day $d+1$.

In both cases, $P$ is true at the end of day $d+1$ and so $P$ must be an invariant. ■

With Lemma 11.6.4 in hand, we can now prove:

**Theorem 11.6.5.** *Everyone is married by the Mating Ritual.*

*Proof.* By contradiction. Assume that it is the last day of the Mating Ritual and someone does not get married. Since there are an equal number of men and women, and since bigamy is not allowed, this means that at least one man (call him Bob) and at least one woman do not get married.

Since Bob is not married, he can't be serenading anybody and so his list must be empty. This means that Bob has crossed every woman off his list and so, by invariant $P$, every woman has a suitor whom she prefers to Bob. Since it is the last day and every woman still has a suitor, this means that every woman gets married. This is a contradiction since we already argued that at least one woman is *not* married. Hence our assumption must be false and so everyone must be married. ■

**Theorem 11.6.6.** *The Mating Ritual produces a stable matching.*

*Proof.* Let Brad and Jen be any man and woman, respectively, that are *not* married to each other on the last day of the Mating Ritual. We will prove that Brad and Jen are not a rogue couple, and thus that all marriages on the last day are stable. There are two cases to consider.

**Case 1:** Jen is not on Brad's list by the end. Then by invariant $P$, we know that Jen has a suitor (and hence a husband) that she prefers to Brad. So she's not going to run off with Brad—Brad and Jen cannot be a rogue couple.

**Case 2:** Jen is on Brad's list. But since Brad is not married to Jen, he must be choosing to serenade his wife instead of Jen, so he must prefer his wife. So he's not going to run off with Jen—once again, Brad and Jen are not a rogue couple. ■

### 11.6.4 . . . Especially the Men

Who is favored by the Mating Ritual, the men or the women? The women *seem* to have all the power: they stand on their balconies choosing the finest among their suitors and spurning the rest. What's more, we know their suitors can only change for the better as the Ritual progresses. Similarly, a man keeps serenading the woman he most prefers among those on his list until he must cross her off, at which point he serenades the next most preferred woman on his list. So from the man's perspective, the woman he is serenading can only change for the worse. Sounds like a good deal for the women.

But it's not! The fact is that from the beginning, the men are serenading their first choice woman, and the desirability of the woman being serenaded decreases only enough to ensure overall stability. The Mating Ritual actually does as well as possible for all the men and does the worst possible job for the women.

To explain all this we need some definitions. Let's begin by observing that while The Mating Ritual produces one stable matching, there may be other stable matchings among the same set of men and women. For example, reversing the roles of men and women will often yield a different stable matching among them.

But some spouses might be out of the question in all possible stable matchings. For example, given the preferences shown in Figure 11.11, Brad is just not in the realm of possibility for Jennifer, since if you ever pair them, Brad and Angelina will form a rogue couple.

**Definition 11.6.7.** Given a set of preference lists for all men and women, one person is in another person's *realm of possible spouses* if there is a stable matching in which the two people are married. A person's *optimal spouse* is their most preferred person within their realm of possibility. A person's *pessimal spouse* is their least preferred person in their realm of possibility.

Everybody has an optimal and a pessimal spouse, since we know there is at least one stable matching, namely, the one produced by the Mating Ritual. Now here is the shocking truth about the Mating Ritual:

**Theorem 11.6.8.** *The Mating Ritual marries every man to his optimal spouse.*

*Proof.* By contradiction. Assume for the purpose of contradiction that some man does not get his optimal spouse. Then there must have been a day when he crossed off his optimal spouse—otherwise he would still be serenading (and would ultimately marry) her or some even more desirable woman.

By the Well Ordering Principle, there must be a *first* day when a man (call him "Keith") crosses off his optimal spouse (call her Nicole). According to the rules of the Ritual, Keith crosses off Nicole because Nicole has a preferred suitor (call him Tom), so

$$\text{Nicole prefers Tom to Keith.} \tag{$*$}$$

Since this is the first day an optimal woman gets crossed off, we know that Tom had not previously crossed off his optimal spouse, and so

$$\text{Tom ranks Nicole at least as high as his optimal spouse.} \tag{$**$}$$

By the definition of an optimal spouse, there must be some stable set of marriages in which Keith gets his optimal spouse, Nicole. But then the preferences given in ($*$) and ($**$) imply that Nicole and Tom are a rogue couple within this supposedly stable set of marriages (think about it). This is a contradiction. ■

**Theorem 11.6.9.** *The Mating Ritual marries every woman to her pessimal spouse.*

*Proof.* Assume for the sake of contradiction that the theorem is not true. Hence there must be a stable set of marriages $\mathcal{M}$ where some woman (call her Nicole) is married to a man (call him Tom) that she likes less than her spouse in The Mating Ritual (call him Keith). This means that

$$\text{Nicole prefers Keith to Tom.} \tag{+}$$

By Theorem 11.6.8 and the fact that Nicole and Keith are married in the Mating Ritual, we know that

$$\text{Keith prefers Nicole to his spouse in } \mathcal{M}. \tag{++}$$

This means that Keith and Nicole form a rogue couple in $\mathcal{M}$, which contradicts the stability of $\mathcal{M}$. ■

### 11.6.5 Applications

The Mating Ritual was first announced in a paper by D. Gale and L.S. Shapley in 1962, but ten years before the Gale-Shapley paper was published, and unknown by them, a similar algorithm was being used to assign residents to hospitals by the National Resident Matching Program (NRMP)[7]. The NRMP has, since the turn of the twentieth century, assigned each year's pool of medical school graduates to hospital residencies (formerly called "internships") with hospitals and graduates playing the roles of men and women. (In this case, there may be multiple women married to one man, a scenario we consider in the problem section at the end of the chapter.). Before the Ritual-like algorithm was adopted, there were chronic disruptions and awkward countermeasures taken to preserve assignments of graduates to residencies. The Ritual resolved these problems so successfully, that it was used essentially without change at least through 1989.[8]

The Internet infrastructure company, Akamai, also uses a variation of the Mating Ritual to assign web traffic to its servers. In the early days, Akamai used other combinatorial optimization algorithms that got to be too slow as the number of servers (over 65,000 in 2010) and requests (over 800 billion per day) increased. Akamai switched to a Ritual-like approach since it is fast and can be run in a distributed

[7] Of course, there is no serenading going on in the hospitals—the preferences are submitted to a program and the whole process is carried out by a computer.

[8] Much more about the Stable Marriage Problem can be found in the very readable mathematical monograph by Dan Gusfield and Robert W. Irving, The Stable Marriage Problem: Structure and Algorithms, MIT Press, Cambridge, Massachusetts, 1989, 240 pp.

manner. In this case, web requests correspond to women and web servers correspond to men. The web requests have preferences based on latency and packet loss, and the web servers have preferences based on cost of bandwidth and colocation.

Not surprisingly, the Mating Ritual is also used by at least one large online dating agency. Even here, there is no serenading going on—everything is handled by computer.

## 11.7 Coloring

In Section 11.2, we used edges to indicate an affinity between a pair of nodes. But there are lots of situations where edges will correspond to *conflicts* between nodes. Exam scheduling is a typical example.

### 11.7.1 An Exam Scheduling Problem

Each term, the MIT Schedules Office must assign a time slot for each final exam. This is not easy, because some students are taking several classes with finals, and (even at MIT) a student can take only one test during a particular time slot. The Schedules Office wants to avoid all conflicts. Of course, you can make such a schedule by having every exam in a different slot, but then you would need hundreds of slots for the hundreds of courses, and the exam period would run all year! So, the Schedules Office would also like to keep exam period short.

The Schedules Office's problem is easy to describe as a graph. There will be a vertex for each course with a final exam, and two vertices will be adjacent exactly when some student is taking both courses. For example, suppose we need to schedule exams for 6.041, 6.042, 6.002, 6.003 and 6.170. The scheduling graph might appear as in Figure 11.13.

6.002 and 6.042 cannot have an exam at the same time since there are students in both courses, so there is an edge between their nodes. On the other hand, 6.042 and 6.170 can have an exam at the same time if they're taught at the same time (which they sometimes are), since no student can be enrolled in both (that is, no student *should* be enrolled in both when they have a timing conflict).

We next identify each time slot with a color. For example, Monday morning is red, Monday afternoon is blue, Tuesday morning is green, etc. Assigning an exam to a time slot is then equivalent to coloring the corresponding vertex. The main constraint is that *adjacent vertices must get different colors*—otherwise, some student has two exams at the same time. Furthermore, in order to keep the exam period short, we should try to color all the vertices using as *few different colors as*

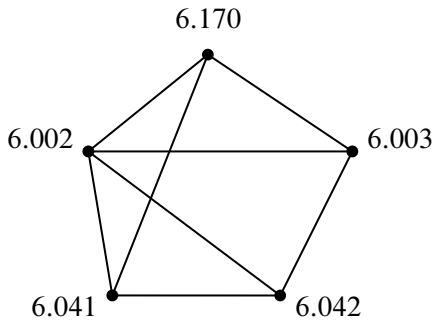



**Figure 11.13** A scheduling graph for five exams. Exams connected by an edge cannot be given at the same time.

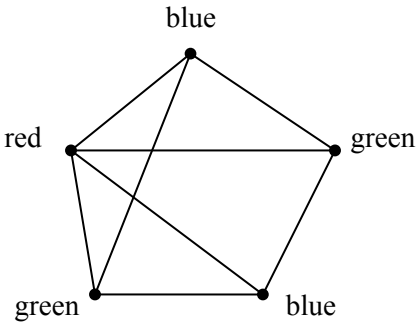



**Figure 11.14** A 3-coloring of the exam graph from Figure 11.13.

*possible*. As shown in Figure 11.14, three colors suffice for our example.

The coloring in Figure 11.14 corresponds to giving one final on Monday morning (red), two Monday afternoon (blue), and two Tuesday morning (green). Can we use fewer than three colors? No! We can't use only two colors since there is a triangle in the graph, and three vertices in a triangle must all have different colors.

This is an example of a *graph coloring* problem: given a graph $G$, assign colors to each node such that adjacent nodes have different colors. A color assignment with this property is called a *valid coloring* of the graph—a "*coloring*," for short. A graph $G$ is *$k$-colorable* if it has a coloring that uses at most $k$ colors.

**Definition 11.7.1.** The minimum value of $k$ for which a graph, $G$, has a valid coloring is called its *chromatic number*, $\chi(G)$.

So $G$ is $k$-colorable iff $\chi(G) \leq k$.

In general, trying to figure out if you can color a graph with a fixed number of colors can take a long time. It's a classic example of a problem for which no fast

algorithms are known. In fact, it is easy to check if a coloring works, but it seems really hard to find it. (If you figure out how, then you can get a $1 million Clay prize.)

### 11.7.2 Some Coloring Bounds

There are some simple properties of graphs that give useful bounds on colorability. The simplest property is being a cycle: an even-length closed cycle is 2-colorable, and since by definition it must have some edges, it is not 1-colorable. So

$$\chi(C_{\text{even}}) = 2.$$

On the other hand, an odd-length cycle requires 3 colors, that is,

$$\chi(C_{\text{odd}}) = 3. \tag{11.3}$$

You should take a moment to think about why this equality holds. Another simple example is a complete graph $K_n$:

$$\chi(K_n) = n$$

since no two vertices can have the same color.

Being bipartite is another property closely related to colorability. If a graph is bipartite, then you can color it with 2 colors using one color for the nodes on the "left" and a second color for the nodes on the "right." Conversely, graphs with chromatic number 2 are all bipartite with all the vertices of one color on the "left" and those with the other color on the right. Since only graphs with no edges —the *empty graphs* —have chromatic number 1, we have:

**Lemma 11.7.2.** *A graph, $G$, with at least one edge is bipartite iff $\chi(G) = 2$.*

The chromatic number of a graph can also be shown to be small if the vertex degrees of the graph are small. In particular, if we have an upper bound on the degrees of all the vertices in a graph, then we can easily find a coloring with only one more color than the degree bound.

**Theorem 11.7.3.** *A graph with maximum degree at most $k$ is $(k+1)$-colorable.*

Since $k$ is the only nonnegative integer valued variable mentioned in the theorem, you might be tempted to try to prove this theorem using induction on $k$. Unfortunately, this approach leads to disaster—we don't know of any reasonable way to do this and expect it would ruin your week if you tried it on a problem set. When you encounter such a disaster using induction on graphs, it is usually best to change what you are inducting on. In graphs, typical good choices for the induction parameter are $n$, the number of nodes, or $e$, the number of edges.

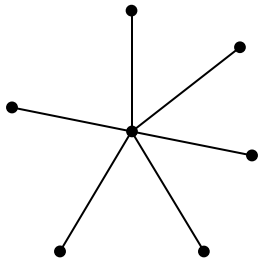


**Figure 11.15** A 7-node star graph.

*Proof of Theorem 11.7.3.* We use induction on the number of vertices in the graph, which we denote by $n$. Let $P(n)$ be the proposition that an $n$-vertex graph with maximum degree at most $k$ is $(k+1)$-colorable.

**Base case** ($n = 1$): A 1-vertex graph has maximum degree 0 and is 1-colorable, so $P(1)$ is true.

**Inductive step**: Now assume that $P(n)$ is true, and let $G$ be an $(n+1)$-vertex graph with maximum degree at most $k$. Remove a vertex $v$ (and all edges incident to it), leaving an $n$-vertex subgraph, $H$. The maximum degree of $H$ is at most $k$, and so $H$ is $(k+1)$-colorable by our assumption $P(n)$. Now add back vertex $v$. We can assign $v$ a color (from the set of $k+1$ colors) that is different from all its adjacent vertices, since there are at most $k$ vertices adjacent to $v$ and so at least one of the $k+1$ colors is still available. Therefore, $G$ is $(k+1)$-colorable. This completes the inductive step, and the theorem follows by induction. ■

Sometimes $k+1$ colors is the best you can do. For example, $\chi(K_n) = n$ and every node in $K_n$ has degree $k = n-1$ and so this is an example where Theorem 11.7.3 gives the best possible bound. By a similar argument, we can show that Theorem 11.7.3 gives the best possible bound for *any* graph with degree bounded by $k$ that has $K_{k+1}$ as a subgraph.

But sometimes $k+1$ colors is far from the best that you can do. For example, the $n$-node *star graph* shown in Figure 11.15 has maximum degree $n-1$ but can be colored using just 2 colors.

### 11.7.3 Why coloring?

One reason coloring problems frequently arise in practice is because scheduling conflicts are so common. For example, at Akamai, a new version of software is deployed over each of 65,000 servers every few days. The updates cannot be done at the same time since the servers need to be taken down in order to deploy the

software. Also, the servers cannot be handled one at a time, since it would take forever to update them all (each one takes about an hour). Moreover, certain pairs of servers cannot be taken down at the same time since they have common critical functions. This problem was eventually solved by making a 65,000-node conflict graph and coloring it with 8 colors—so only 8 waves of install are needed!

Another example comes from the need to assign frequencies to radio stations. If two stations have an overlap in their broadcast area, they can't be given the same frequency. Frequencies are precious and expensive, so you want to minimize the number handed out. This amounts to finding the minimum coloring for a graph whose vertices are the stations and whose edges connect stations with overlapping areas.

Coloring also comes up in allocating registers for program variables. While a variable is in use, its value needs to be saved in a register. Registers can be reused for different variables but two variables need different registers if they are referenced during overlapping intervals of program execution. So register allocation is the coloring problem for a graph whose vertices are the variables: vertices are adjacent if their intervals overlap, and the colors are registers. Once again, the goal is to minimize the number of colors needed to color the graph.

Finally, there's the famous map coloring problem stated in Proposition 1.1.6. The question is how many colors are needed to color a map so that adjacent territories get different colors? This is the same as the number of colors needed to color a graph that can be drawn in the plane without edges crossing. A proof that four colors are enough for *planar* graphs was acclaimed when it was discovered about thirty years ago. Implicit in that proof was a 4-coloring procedure that takes time proportional to the number of vertices in the graph (countries in the map).

Surprisingly, it's another of those million dollar prize questions to find an efficient procedure to tell if a planar graph really *needs* four colors, or if three will actually do the job. A proof that testing 3-colorability of graphs is as hard as the million dollar SAT problem is given in Problem 11.24; this turns out to be true even for planar graphs. (It is easy to tell if a graph is 2-colorable, as explained in Section 11.10.) In Chapter 12, we'll develop enough planar graph theory to present an easy proof that all planar graphs are 5-colorable.

## 11.8 Getting from *u* to *v* in a Graph

Walks and paths in simple graphs are esentially the same as in digraphs. We just modify the digraph definitions using undirected edges instead of directed ones. For example, the formal definition of a walk in a simple graph is a virtually that same

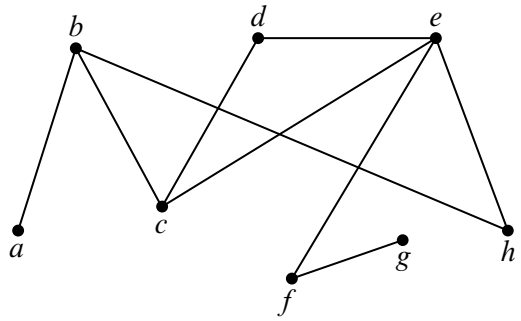



**Figure 11.16** *A graph with 3 cycles: bhecb, cdec, bcdehb.*

as the Definition 9.2.1 of a walk in a digraph:

**Definition 11.8.1.** A *walk in a simple graph*, $G$, is an alternating sequence of vertices and edges that begins with a vertex, ends with a vertex, and such that for every edge $\langle u—v\rangle$ in the walk, one of the endpoints $u$, $v$ is the element just before the edge, and the other endpoint is the next element after the edge. The *length of a walk* is the total number of occurrences of edges in it.

So a walk, $\mathbf{v}$, is a sequence of the form

$$\mathbf{v} ::= v_0\ \langle v_0—v_1\rangle\ v_1\ \langle v_1—v_2\rangle\ v_2\ \ldots\ \langle v_{k-1}—v_k\rangle\ v_k$$

where $\langle v_i—v_{i+1}\rangle \in E(G)$ for $i \in [0,k)$. The walk is said to *start* at $v_0$, to *end* at $v_k$, and the *length*, $|\mathbf{v}|$, of the walk is $k$. The walk is a *path* iff all the $v_i$'s are different, that is, if $i \neq j$, then $v_i \neq v_j$.

A *closed walk* is a walk that begins and ends at the same vertex. A *cycle* is a closed walk of length three or more whose vertices are distinct except for the beginning and end vertices.

Note that a single vertex counts as a length zero path and closed walk. But in contrast to digraphs, a single vertex is not considered to be a cycle.

As in digraphs, the length of a walk is *one less* than the number of occurrences of vertices in it. For example, the graph in Figure 11.16 has a length 6 path through the seven successive vertices $abcdefg$. This is the longest path in the graph. The graph in Figure 11.16 also has three cycles through successive vertices $bhecb$, $cdec$, and $bcdehb$.

### 11.8.1 Cycles as Subgraphs

A cycle does not really have a beginning or an end, and so can be described by *any* of the paths that go around it. For example, in the graph in Figure 11.16, the cycle

starting at $b$ and going through vertices $bcdehb$ can also be described as starting at $d$ and going through $decbcd$. Furthermore, cycles in simple graphs don't have a direction: $dcbced$ describes the same cycle as though it started and ended at $d$ but went in the opposite direction.

A precise way to explain which closed walks describe the same cycle is to define cycle as a subgraph instead of as a closed walk. Namely, we could define a cycle in $G$ to be a *subgraph* of $G$ that looks like a length-$n$ cycle for $n \geq 3$.

**Definition 11.8.2.** A graph $G$ is said to be a *subgraph* of a graph $H$ if $V(G) \subseteq V(H)$ and $E(G) \subseteq E(H)$.

For example, the one-edge graph $G$ where

$$V(G) = \{g, h, i\} \quad \text{and} \quad E(G) = \{\langle h\text{—}i\rangle\}$$

is a subgraph of the graph $H$ in Figure 11.1. On the other hand, any graph containing an edge $\langle g\text{—}h\rangle$ will not be a subgraph of $H$ because this edge is not in $E(H)$. Another example is an empty graph on $n$ nodes, which will be a subgraph of an $L_n$ with same set of nodes; similarly, $L_n$ is a subgraph of $C_n$, and $C_n$ is a subgraph of $K_n$.

**Definition 11.8.3.** For $n \geq 3$, let $C_n$ be the graph with vertices $1, \ldots, n$ and edges

$$\langle 1\text{—}2\rangle\,,\ \langle 2\text{—}3\rangle\,,\ \ldots,\ \langle (n-1)\text{—}n\rangle\,,\ \langle n\text{—}1\rangle\,.$$

A *cycle of a graph*, $G$, is a subgraph of $G$ that is isomorphic to $C_n$ for some $n \geq 3$.

This definition formally captures the idea that cycles don't have direction or beginnings or ends.

## 11.9 Connectivity

**Definition 11.9.1.** *Two vertices are connected* in a graph when there is a path that begins at one and ends at the other. By convention, every vertex is connected to itself by a path of length zero. A *graph is connected* when every pair of vertices are connected.

### 11.9.1 Connected Components

Being connected is usually a good property for a graph to have. For example, it could mean that it is possible to get from any node to any other node, or that it is possible to communicate between any pair of nodes, depending on the application.

But not all graphs are connected. For example, the graph where nodes represent cities and edges represent highways might be connected for North American cities, but would surely not be connected if you also included cities in Australia. The same is true for communication networks like the Internet—in order to be protected from viruses that spread on the Internet, some government networks are completely isolated from the Internet.

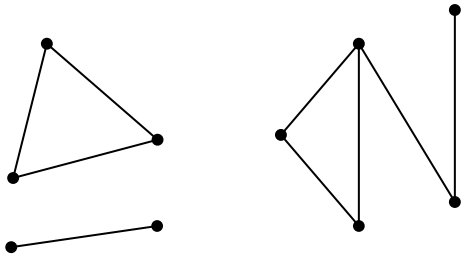


**Figure 11.17** One graph with 3 connected components.

Another example, is shown in Figure 11.17, which looks like a picture of three graphs, but is intended to be a picture of *one* graph. This graph consists of three pieces (subgraphs). Each piece by itself is connected, but there are no paths between vertices in different pieces. These connected pieces of a graph are called its *connected components*.

**Definition 11.9.2.** A *connected component* of a graph is a subgraph consisting of some vertex and every node and edge that is connected to that vertex.

So a graph is connected iff it has exactly one connected component. At the other extreme, the empty graph on $n$ vertices has $n$ connected components.

### 11.9.2 $k$-Connected Graphs

If we think of a graph as modeling cables in a telephone network, or oil pipelines, or electrical power lines, then we not only want connectivity, but we want connectivity that survives component failure. So more generally we want to define how strongly two vertices are connected. One measure of connection strength is how many links must fail before connectedness fails. In particular, two vertices are *$k$-edge connected* when it takes at least $k$ "edge-failures" to disconnect them. More precisely:

**Definition 11.9.3.** Two vertices in a graph are *$k$-edge connected* when they remain connected in every subgraph obtained by deleting up to $k-1$ edges. A graph is $k$-edge connected when it has more than one vertex, and every subgraph obtained by deleting at most $k-1$ edges is connected.

So two vertices are connected according to Definition 11.9.1 iff they are 1-edge connected according to Definition 11.9.3; likewise for any graph with more than one vertex.

There are other kinds of connectedness but edge-connectedness will be enough for us, so from now on we'll drop the "edge" modifier and just say "connected."[9]

For example, in the graph in Figure 11.16, vertices $c$ and $e$ are 3 connected, $b$ and $e$ are 2 connected, $g$ and $e$ are 1 connected, and no vertices are 4 connected. The graph as a whole is only 1 connected. A complete graph, $K_n$, is $(n-1)$ connected. Every cycle is 2-connected.

The idea of a *cut edge* is a useful way to explain 2-connectivity.

**Definition 11.9.4.** If two vertices are connected in a graph $G$, but not connected when an edge $e$ is removed, then $e$ is called a *cut edge* of $G$.

So a graph with more than one vertex is 2-connected iff it is connected, and has no cut edges. The following Lemma is another immediate consequence of the definition:

**Lemma 11.9.5.** *An edge is a cut edge iff it is not on a cycle.*

More generally, if two vertices are connected by $k$ edge-disjoint paths —that is, no edge occurs in two paths —then they must be $k$ connected, since at least one edge will have to be removed from each of the paths before they could disconnect. A fundamental fact, whose ingenious proof we omit, is Menger's theorem which confirms that the converse is also true: if two vertices are $k$-connected, then there are $k$ edge-disjoint paths connecting them. It takes some ingenuity to prove this just for the case $k = 2$.

### 11.9.3 The Minimum Number of Edges in a Connected Graph

The following theorem says that a graph with few edges must have many connected components.

**Theorem 11.9.6.** *Every graph, $G$, has at least $|V(G)| - |E(G)|$ connected components.*

[9]There is an obvious definition of $k$-vertex connectedness based on deleting vertices rather than edges. Graph theory texts usually use "$k$-connected" as shorthand for "$k$-vertex connected."

Of course for Theorem 11.9.6 to be of any use, there must be fewer edges than vertices.

*Proof.* We use induction on the number, $k$, of edges. Let $P(k)$ be the proposition that

> every graph, $G$, with $k$ edges has at least $|V(G)| - k$ connected components.

**Base case** ($k = 0$): In a graph with 0 edges, each vertex is itself a connected component, and so there are exactly $|V(G)| = |V(G)| - 0$ connected components. So $P(0)$ holds.

**Inductive step**:

Let $G_e$ be the graph that results from removing an edge, $e \in E(G)$. So $G_e$ has $k$ edges, and by the induction hypothesis $P(k)$, we may assume that $G_e$ has at least $|V(G)| - k$ connected components. Now add back the edge $e$ to obtain the original graph $G$. If the endpoints of $e$ were in the same connected component of $G_e$, then $G$ has the same sets of connected vertices as $G_e$, so $G$ has at least $|V(G)| - k > |V(G)| - (k+1)$ components. Alternatively, if the endpoints of $e$ were in different connected components of $G_e$, then these two components are merged into one component in $G$, while all other components remain unchanged, so that $G$ has one fewer connected component than $G_e$. That is, $G$ has at least $(|V(G)| - k) - 1 = |V(G)| - (k+1)$ connected components. So in either case, $G$ has at least $|V(G)| - (k+1)$ components, as claimed.

This completes the inductive step and hence the entire proof by induction. ■

**Corollary 11.9.7.** *Every connected graph with $n$ vertices has at least $n - 1$ edges.*

A couple of points about the proof of Theorem 11.9.6 are worth noticing. First, we used induction on the number of edges in the graph. This is very common in proofs involving graphs, as is induction on the number of vertices. When you're presented with a graph problem, these two approaches should be among the first you consider.

The second point is more subtle. Notice that in the inductive step, we took an arbitrary $(k+1)$-edge graph, threw out an edge so that we could apply the induction assumption, and then put the edge back. You'll see this shrink-down, grow-back process very often in the inductive steps of proofs related to graphs. This might seem like needless effort: why not start with an $k$-edge graph and add one more to get an $(k+1)$-edge graph? That would work fine in this case, but opens the door to a nasty logical error called *buildup error* illustrated in Problem 11.29.

## 11.10 Odd Cycles and 2-Colorability

We have already seen that determining the chromatic number of a graph is a challenging problem. There is one special case where this problem is very easy, namely, when the graph is 2-colorable.

**Theorem 11.10.1.** *The following graph properties are equivalent:*

1. *The graph contains an odd length cycle.*
2. *The graph is not 2-colorable.*
3. *The graph contains an odd length closed walk.*

In other words, if a graph has any one of the three properties above, then it has all of the properties.

We will show the following implications among these properties:

1. IMPLIES 2. IMPLIES 3. IMPLIES 1.

So each of these properties implies the other two, which means they all are equivalent.

**1 IMPLIES 2** *Proof.* This follows from equation 11.3. ■

**2 IMPLIES 3** If we prove this implication for connected graphs, then it will hold for an arbitrary graph because it will hold for each connected component. So we can assume that $G$ is connected.

*Proof.* Pick an arbitrary vertex $r$ of $G$. Since $G$ is connected, for every node $u \in V(G)$, there will be a walk $\mathbf{w}_u$ starting at $u$ and ending at $r$. Assign colors to vertices of $G$ as follows:

$$\text{color}(u) = \begin{cases} \text{black}, & \text{if } |\mathbf{w}_u| \text{ is even,} \\ \text{white}, & \text{otherwise.} \end{cases}$$

Now since $G$ is not colorable, this can't be a valid coloring. So there must be an edge between two nodes $u$ and $v$ with the same color. But in that case

$$\mathbf{w}_u \widehat{\ } \text{reverse}(\mathbf{w}_v) \widehat{\ } \langle v\text{—}u \rangle$$

is a closed walk starting and ending at $u$, and its length is

$$|\mathbf{w}_u| + |\mathbf{w}_v| + 1.$$

This length is odd, since $\mathbf{w}_u$ and $\mathbf{w}_v$ are both even length or are both odd length. ■

**3 IMPLIES 1** *Proof.* Since there is an odd length closed walk, the WOP implies there is a odd length closed walk $\mathbf{w}$ of minimum length. We claim $\mathbf{w}$ must be a cycle. To show this, assume to the contrary that there is vertex $x$ that appears twice on the walk, so $\mathbf{w}$ consists of a closed walk from $x$ to $x$ followed by another such walk. That is,

$$\mathbf{w} = \mathbf{f}\,\widehat{x}\,\mathbf{r}$$

for some positive length walks $\mathbf{f}$ and $\mathbf{r}$ that begin and end at $x$. Since

$$|\mathbf{w}| = |\mathbf{f}| + |\mathbf{r}|$$

is odd, exactly one of $\mathbf{f}$ and $\mathbf{g}$ must have odd length, and that one will be an odd length closed walk shorter than $\mathbf{w}$, a contradiction.

■

This completes the proof of Theorem 11.10.1.

Theorem 11.10.1 turns out to be useful since bipartite graphs come up fairly often in practice. We'll see examples when we talk about planar graphs in Chapter 12.

## 11.11 Forests & Trees

We've already made good use of digraphs without cycles, but *simple* graphs without cycles are arguably the most important graphs of all in computer science.

### 11.11.1 Leaves, Parents & Children

**Definition 11.11.1.** An acyclic graph is called a *forest*. A connected acyclic graph is called a *tree*.

The graph shown in Figure 11.18 is a forest. Each of its connected components is by definition a tree.

One of the first things you will notice about trees is that they tend to have a lot of nodes with degree one. Such nodes are called *leaves*.

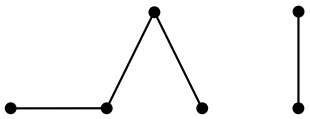


**Figure 11.18** A 6-node forest consisting of 2 component trees.

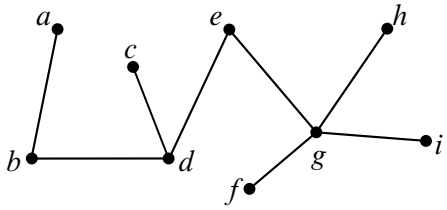



**Figure 11.19** A 9-node tree with 5 leaves.

**Definition 11.11.2.** A degree 1 node in a forest is called a *leaf*.

The forest in Figure 11.18 has 4 leaves. The tree in Figure 11.19 has 5 leaves.

Trees are a fundamental data structure in computer science. For example, information is often stored in tree-like data structures and the execution of many recursive programs can be modeled as the traversal of a tree. In such cases, it is often useful to arrange the nodes in levels, where the node at the top level is identified as the *root* and where every edge joins a *parent* to a *child* one level below. Figure 11.20 shows the tree of Figure 11.19 redrawn in this way. Node $d$ is a child of node $e$ and the parent of nodes $b$ and $c$.

### 11.11.2 Properties

Trees have many unique properties. We have listed some of them in the following theorem.

**Theorem 11.11.3.** *Every tree has the following properties:*

1. *Every connected subgraph is a tree.*
2. *There is a unique path between every pair of vertices.*
3. *Adding an edge between nonadjacent nodes in a tree creates a graph with a cycle.*
4. *Removing any edge disconnects the graph. That is, every edge is a cut edge.*
5. *If the tree has at least two vertices, then it has at least two leaves.*

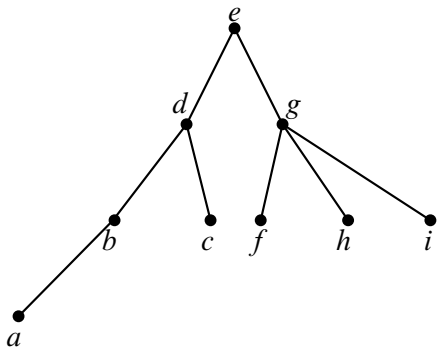



**Figure 11.20** The tree from Figure 11.19 redrawn with node $e$ as the root and the other nodes arranged in levels.

6. *The number of vertices in a tree is one larger than the number of edges.*

*Proof.* 1. A cycle in a subgraph is also a cycle in the whole graph, so any subgraph of an acyclic graph must also be acyclic. If the subgraph is also connected, then by definition, it is a tree.

2. Since a tree is connected, there is at least one path between every pair of vertices. Suppose for the purposes of contradiction, that there are two different paths between some pair of vertices. Then there are two distinct paths $\mathbf{p} \neq \mathbf{q}$ between the same two vertices with minimum total length $|\mathbf{p}| + |\mathbf{q}|$. If these paths shared a vertex, $w$, other than at the start and end of the paths, then the parts of $\mathbf{p}$ and $\mathbf{q}$ from start to $w$, or the parts of $\mathbf{p}$ and $\mathbf{q}$ from $w$ to the end, must be distinct paths between the same vertices with total length less than $|\mathbf{p}| + |\mathbf{q}|$, contradicting the minimality of this sum. Therefore, $\mathbf{p}$ and $\mathbf{q}$ have no vertices in common besides their endpoints, and so $\mathbf{p}^\frown\text{reverse}(\mathbf{q})$ is a cycle.

3. An additional edge $\langle u\text{—}v\rangle$ together with the unique path between $u$ and $v$ forms a cycle.

4. Suppose that we remove edge $\langle u\text{—}v\rangle$. Since the tree contained a unique path between $u$ and $v$, that path must have been $\langle u\text{—}v\rangle$. Therefore, when that edge is removed, no path remains, and so the graph is not connected.

5. Since the tree has at least two vertices, the longest path in the tree will have different endpoints $u$ and $v$. We claim $u$ is a leaf. This follows because, since by definition of endpoint, $u$ is incident to at most one edge on the path.

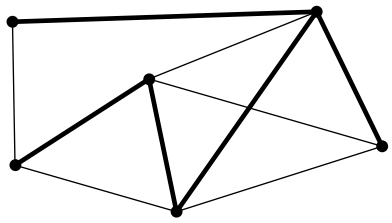


**Figure 11.21** A graph where the edges of a spanning tree have been thickened.

Also, If $u$ was incident to an edge not on the path, then the path could be lengthened by adding that edge, contradicting the fact that the path was as long as possible. It follows that $u$ is incident only to a single edge, that is $u$ is a leaf. The same hold for $v$.

6. We use induction on the proposition

$$P(n) ::= \text{there are } n-1 \text{ edges in any } n\text{-vertex tree.}$$

**Base case** ($n = 1$): $P(1)$ is true since a tree with 1 node has 0 edges and $1-1=0$.

**Inductive step**: Now suppose that $P(n)$ is true and consider an $(n+1)$-vertex tree, $T$. Let $v$ be a leaf of the tree. You can verify that deleting a vertex of degree 1 (and its incident edge) from any connected graph leaves a connected subgraph. So by Theorem 11.11.3.1, deleting $v$ and its incident edge gives a smaller tree, and this smaller tree has $n-1$ edges by induction. If we re-attach the vertex, $v$, and its incident edge, we find that $T$ has $n = (n+1)-1$ edges. Hence, $P(n+1)$ is true, and the induction proof is complete. ■

Various subsets of properties in Theorem 11.11.3 provide alternative characterizations of trees. For example,

**Lemma 11.11.4.** *A graph $G$ is a tree iff $G$ is a forest and $|V(G)| = |E(G)| + 1$.*

The proof is an easy consequence of Theorem 11.9.6.6.

### 11.11.3 Spanning Trees

Trees are everywhere. In fact, every connected graph contains a subgraph that is a tree with the same vertices as the graph. This is called a *spanning tree* for the graph. For example, Figure 11.21 is a connected graph with a spanning tree highlighted.

**Definition 11.11.5.** Define a *spanning subgraph* of a graph, $G$, to be a subgraph containing all the vertices of $G$.

**Theorem 11.11.6.** *Every connected graph contains a spanning tree.*

*Proof.* Suppose $G$ is a connected graph, so the graph $G$ itself is a connected, spanning subgraph. So by WOP, $G$ must have a minimum-edge connected, spanning subgraph, $T$. We claim $T$ is a spanning tree. Since $T$ is a connected, spanning subgraph by definition, all we have to show is that $T$ is acyclic.

But suppose to the contrary that $T$ contained a cycle $C$. By Lemma 11.9.5, an edge $e$ of $C$ will not be a cut edge, so removing it would leave a connected, spanning subgraph that was smaller than $T$, contradicting the minimality to $T$. ■

### 11.11.4 Minimum Weight Spanning Trees

Spanning trees are interesting because they connect all the nodes of a graph using the smallest possible number of edges. For example the spanning tree for the 6-node graph shown in Figure 11.21 has 5 edges.

Spanning trees are very useful in practice, but in the real world, not all spanning trees are equally desirable. That's because, in practice, there are often costs associated with the edges of the graph.

For example, suppose the nodes of a graph represent buildings or towns and edges represent connections between buildings or towns. The cost to actually make a connection may vary a lot from one pair of buildings or towns to another. The cost might depend on distance or topography. For example, the cost to connect LA to NY might be much higher than that to connect NY to Boston. Or the cost of a pipe through Manhattan might be more than the cost of a pipe through a cornfield.

In any case, we typically represent the cost to connect pairs of nodes with a weighted edge, where the weight of the edge is its cost. The weight of a spanning tree is then just the sum of the weights of the edges in the tree. For example, the weight of the spanning tree shown in Figure 11.22 is 19.

The goal, of course, is to find the spanning tree with minimum weight, called the minimum weight spanning tree (MST for short).

**Definition 11.11.7.** A *minimum weight spanning tree* (MST) of an edge-weighted graph $G$ is a spanning tree of $G$ with the smallest possible sum of edge weights.

Is the spanning tree shown in Figure 11.22(a) an MST of the weighted graph shown in Figure 11.22(b)? Actually, it is not, since the tree shown in Figure 11.23 is also a spanning tree of the graph shown in Figure 11.22(b), and this spanning tree has weight 17.

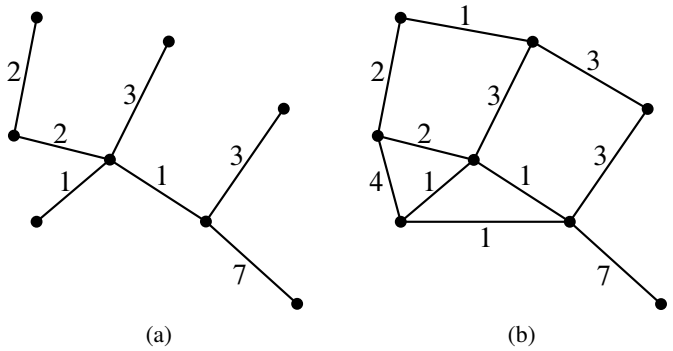



**Figure 11.22** A spanning tree (a) with weight 19 for a graph (b).

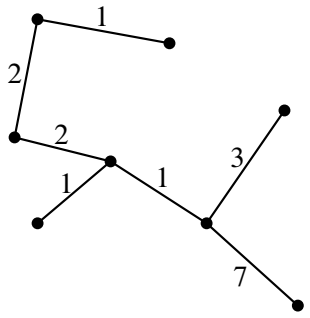



**Figure 11.23** An MST with weight 17 for the graph in Figure 11.22(b).

What about the tree shown in Figure 11.23? Is it an MST? It seems to be, but how do we prove it? In general, how do we find an MST for a connected graph $G$? We could try enumerating all subtrees of $G$, but that approach would be hopeless for large graphs.

There actually are many good ways to find MST's based on an invariance property of some subgraphs of $G$ called pre-MST's.

**Definition 11.11.8.** A *pre-MST* for a graph $G$ is a spanning subgraph of $G$ that is also a subgraph of some MST of $G$.

So a pre-MST will necessarily be a forest.

For example, the empty graph with the same vertices as $G$ is guaranteed to be a pre-MST of $G$, and so is any actual MST of $G$.

If $e$ is an edge of $G$ and $S$ is a spanning subgraph, we'll write $S + e$ for the spanning subgraph with edges $E(S) \cup \{e\}$.

**Definition 11.11.9.** If $F$ is a pre-MST and $e$ is a new edge, that is $e \in E(G) - E(F)$, then $e$ *extends* $F$ when $F + e$ is also a pre-MST.

So being a pre-MST is by definition an invariant under addition of extending edges.

The standard methods for finding MST's all start with the empty spanning forest and build up to an MST by adding one extending edge after another. Since the empty spanning forest is a pre-MST, and being a pre-MST is invariant under extensions, every forest built in this way will be a pre-MST. But no spanning tree can be a subgraph of a different spanning tree. So when the pre-MST finally grows enough to become a tree, it will be an MST. By Lemma 11.11.4, this happens after exactly $|V(G)| - 1$ edge extensions.

So the problem of finding MST's reduces to the question of how to tell if an edge is an extending edge. Here's how:

**Definition 11.11.10.** Let $F$ be a pre-MST, and color the vertices in each connected component of $F$ either all black or all white. At least one component of each color is required. Call this a *solid coloring* of $F$. A *gray edge* of a solid coloring is an edge of $G$ with different colored endpoints.

Any path in $G$ from a white vertex to a black vertex obviously must include a gray edge, so for any solid coloring, there is guaranteed to be at least one gray edge. In fact, there will have to be at least as many gray edges as there are components with the same color. Here's the punchline:

**Lemma 11.11.11.** *An edge extends a pre-MST $F$ if it is a minimum weight gray edge in some solid coloring of $F$.*

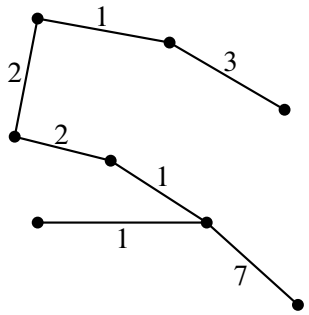



**Figure 11.24** A spanning tree found by Algorithm 1.

So to extend a pre-MST, choose any solid coloring, find the gray edges, and among them choose one with minimum weight. Each of these steps is easy to do, so it is easy to keep extending and arrive at an MST. For example, here are three known algorithms that are explained by Lemma 11.11.11:

**Algorithm 1.** *[Prim] Grow a tree one edge at a time by adding a minimum weight edge among the edges that have exactly one endpoint in the tree.*

This is the algorithm that comes from coloring the growing tree white and all the vertices not in the tree black. Then the gray edges are the ones with exactly one endpoint in the tree.

**Algorithm 2.** *[Kruskal] Grow a forest one edge at a time by adding a minimum weight edge among the edges with endpoints in different connected components.*

The edges between different components are exactly the edges that are gray under some solid coloring, namely any coloring where the components it connects have different colors.

For example, in the weighted graph we have been considering, we might run Algorithm 1 as follows. We would start by choosing one of the weight 1 edges, since this is the smallest weight in the graph. Suppose we chose the weight 1 edge on the bottom of the triangle of weight 1 edges in our graph. This edge is incident to the same vertex as two weight 1 edges, a weight 4 edge, a weight 7 edge, and a weight 3 edge. We would then choose the incident edge of minimum weight. In this case, one of the two weight 1 edges. At this point, we cannot choose the third weight 1 edge: it won't be gray because its endpoints are both in the tree, and so are both colored white. But we can continue by choosing a weight 2 edge. We might end up with the spanning tree shown in Figure 11.24, which has weight 17, the smallest we've seen so far.

Now suppose we instead ran Algorithm 2 on our graph. We might again choose the weight 1 edge on the bottom of the triangle of weight 1 edges in our graph. Now, instead of choosing one of the weight 1 edges it touches, we might choose the weight 1 edge on the top of the graph. This edge still has minimum weight, and will be gray if we simply color its endpoints differently, so Algorithm 2 can choose it. We would then choose one of the remaining weight 1 edges. Note that neither causes us to form a cycle. Continuing the algorithm, we could end up with the same spanning tree in Figure 11.24, though this will depend on how the tie breaking rules used to choose among gray edges with the same minimum weight. For example, if the weight of every edge in $G$ is one, then all spanning trees are MST's with weight $|V(G)| - 1$, and both of these algorithms can arrive at each of these spanning trees by suitable tie-breaking.

The coloring that explains Algorithm 1 also justifies a more flexible algorithm which has Algorithm 1 as a special case:

**Algorithm 3.** *Grow a forest one edge at a time by picking any component and adding a minimum weight edge among the edges leaving that component.*

This algorithm allows components that are not too close to grow in parallel and independently, which is great for "distributed" computation where separate processors share the work with limited communication between processors.

These are examples of greedy approaches to optimization. Sometimes greediness works and sometimes it doesn't. The good news is that it does work to find the MST. So we can be sure that the MST for our example graph has weight 17 since it was produced by Algorithm 2. And we have a fast algorithm for finding a minimum weight spanning tree for any graph.

Ok, to wrap up this story, all that's left is the proof that minimal gray edges are extending edges. This might sound like a chore, but it just uses the same reasoning we used to be sure there would be a gray edge when you need it.

*Proof.* (of Lemma 11.11.11)

Let $F$ be a pre-MST that is a subgraph of some MST $M$ of $G$, and suppose $e$ is a minimum weight gray edge under some solid coloring of $F$. We want to show that $F + e$ is also a pre-MST.

If $e$ happens to be an edge of $M$, then $F + e$ remains a subgraph of $M$, and so is a pre-MST.

The other case is when $e$ is not an edge of $M$. In that case, $M + e$ will be a connected, spanning subgraph. Also $M$ has a path $\mathbf{p}$ between the different colored endpoints of $e$, so $M + e$ has a cycle consisting of $e$ together with $\mathbf{p}$. Now $\mathbf{p}$ has both a black endpoint and a white one, so it must contain some gray edge $g \neq e$. The trick is to remove $g$ from $M + e$ to obtain a subgraph $M + e - g$. Since gray

edges by definition are not edges of $F$, the graph $M + e - g$ contains $F + e$. We claim that $M + e - g$ is an MST, which proves the claim that $e$ extends $F$.

To prove this claim, note that $M + e$ is a connected, spanning subgraph, and $g$ is on a cycle of $M + e$, so by Lemma 11.9.5, removing $g$ won't disconnect anything. Therefore, $M + e - g$ is still a connected, spanning subgraph. Moreover, $M + e - g$ has the same number of edges as $M$, so Lemma 11.11.4 implies that it must be a spanning tree. Finally, since $e$ is minimum weight among gray edges,

$$w(M + e - g) = w(M) + w(e) - w(g) \leq w(M).$$

This means that $M + e - g$ is a spanning tree whose weight is at most that of an MST, which implies that $M + e - g$ is also an MST. ■

Another interesting fact falls out of the proof of Lemma 11.11.11:

**Corollary 11.11.12.** *If all edges in a weighted graph have distinct weights, then the graph has a* unique *MST.*

The proof of Corollary 11.11.12 is left to Problem 11.42.

## Problems for Section 11.2

### Class Problems

**Problem 11.1. (a)** Prove that in every graph, there are an even number of vertices of odd degree.

*Hint:* The Handshaking Lemma 11.2.1.

**(b)** Conclude that at a party where some people shake hands, the number of people who shake hands an odd number of times is an even number.

**(c)** Call a sequence of two or more different people at the party a *handshake sequence* if, except for the last person, each person in the sequence has shaken hands with the next person in the sequence.

Suppose George was at the party and has shaken hands with an odd number of people. Explain why, starting with George, there must be a handshake sequence ending with a different person who has shaken an odd number of hands.

*Hint:* Just look at the people at the ends of handshake sequences that start with George.

### Exam Problems

**Problem 11.2.**
A researcher analyzing data on heterosexual sexual behavior in a group of $m$ males and $f$ females found that within the group, the male average number of female partners was 10% larger that the female average number of male partners.

**(a)** Comment on the following claim. "Since we're assuming that each encounter involves one man and one woman, the average numbers should be the same, so the males must be exaggerating."

**(b)** For what constant $c$ is $m = c \cdot f$?

**(c)** The data shows that approximately 20% of the females were virgins, while only 5% of the males were. The researcher wonders how excluding virgins from the population would change the averages. If he knew graph theory, the researcher would realize that the nonvirgin male average number of partners will be $x(f/m)$ times the nonvirgin female average number of partners. What is $x$?

**(d)** For purposes of further research, it would be helpful to pair each female in the group with a unique male in the group. Explain why this is not possible.

## Problems for Section 11.4

### Class Problems

**Problem 11.3.**
For each of the following pairs of graphs, either define an isomorphism between them, or prove that there is none. (We write $ab$ as shorthand for $\langle a\text{—}b\rangle$.)

**(a)**

$$G_1 \text{ with } V_1 = \{1,2,3,4,5,6\},\ E_1 = \{12,23,34,14,15,35,45\}$$
$$G_2 \text{ with } V_2 = \{1,2,3,4,5,6\},\ E_2 = \{12,23,34,45,51,24,25\}$$

**(b)**

$$G_3 \text{ with } V_3 = \{1,2,3,4,5,6\},\ E_3 = \{12,23,34,14,45,56,26\}$$
$$G_4 \text{ with } V_4 = \{a,b,c,d,e,f\},\ E_4 = \{ab,bc,cd,de,ae,ef,cf\}$$

### Homework Problems

**Problem 11.4.**
Determine which among the four graphs pictured in the Figure 11.25 are isomorphic. If two of these graphs are isomorphic, describe an isomorphism between

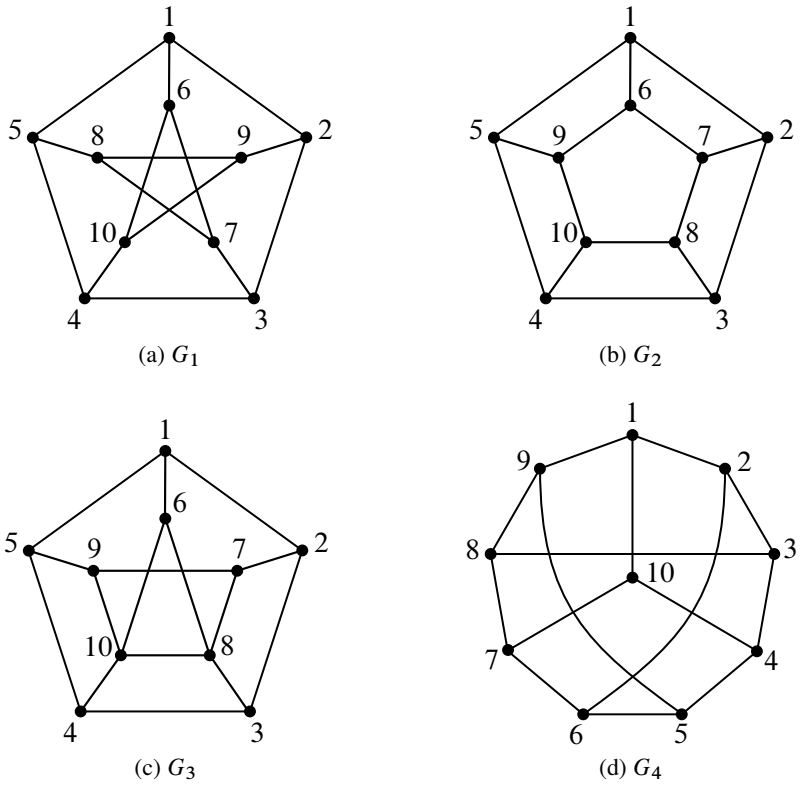



**Figure 11.25** Which graphs are isomorphic?

them. If they are not, give a property that is preserved under isomorphism such that one graph has the property, but the other does not. For at least one of the properties you choose, *prove* that it is indeed preserved under isomorphism (you only need prove one of them).

**Problem 11.5. (a)** For any vertex, $v$, in a graph, let $N(v)$ be the set of *neighbors* of $v$, namely, the vertices adjacent to $v$:

$$N(v) ::= \{u \mid \langle u\text{—}v\rangle \text{ is an edge of the graph}\}.$$

Suppose $f$ is an isomorphism from graph $G$ to graph $H$. Prove that $f(N(v)) = N(f(v))$.

Your proof should follow by simple reasoning using the definitions of isomorphism and neighbors—no pictures or handwaving.

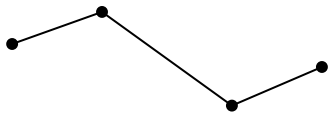
*Hint:* Prove by a chain of iff's that

$$h \in N(f(v)) \quad \text{iff} \quad h \in f(N(v))$$

for every $h \in V_H$. Use the fact that $h = f(u)$ for some $u \in V_G$.

**(b)** Conclude that if $G$ and $H$ are isomorphic graphs, then for each $k \in \mathbb{N}$, they have the same number of degree $k$ vertices.

**Problem 11.6.**
Let's say that a graph has "two ends" if it has exactly two vertices of degree 1 and all its other vertices have degree 2. For example, here is one such graph:



**(a)** A *line graph* is a graph whose vertices can be listed in a sequence with edges between consecutive vertices only. So the two-ended graph above is also a line graph of length 4.

Prove that the following theorem is false by drawing a counterexample.
**False Theorem.** *Every two-ended graph is a line graph.*

**(b)** Point out the first erroneous statement in the following bogus proof of the false theorem and describe the error.

*Bogus proof.* We use induction. The induction hypothesis is that every two-ended graph with $n$ edges is a path.

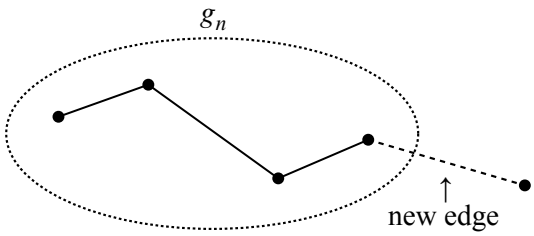
**Base case ($n = 1$):** The only two-ended graph with a single edge consists of two vertices joined by an edge:



Sure enough, this is a line graph.

**Inductive case:** We assume that the induction hypothesis holds for some $n \geq 1$ and prove that it holds for $n + 1$. Let $G_n$ be any two-ended graph with $n$ edges. By the induction assumption, $G_n$ is a line graph. Now suppose that we create a

two-ended graph $G_{n+1}$ by adding one more edge to $G_n$. This can be done in only one way: the new edge must join an endpoint of $G_n$ to a new vertex; otherwise, $G_{n+1}$ would not be two-ended.

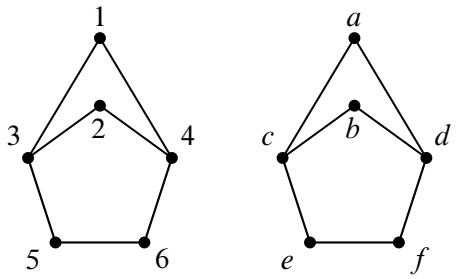



Clearly, $G_{n+1}$ is also a line graph. Therefore, the induction hypothesis holds for all graphs with $n + 1$ edges, which completes the proof by induction.

■

### Exam Problems

**Problem 11.7.**
There are four isomorphisms between these two graphs. List them.




## Problems for Section 11.5

### Class Problems

**Problem 11.8.**
A certain Institute of Technology has a lot of student clubs; these are loosely overseen by the Student Association. Each eligible club would like to delegate one of its members to appeal to the Dean for funding, but the Dean will not allow a student to be the delegate of more than one club. Fortunately, the Association VP took Math for Computer Science and recognizes a matching problem when she sees one.

**(a)** Explain how to model the delegate selection problem as a bipartite matching problem.

**(b)** The VP's records show that no student is a member of more than 9 clubs. The VP also knows that to be eligible for support from the Dean's office, a club must have at least 13 members. That's enough for her to guarantee there is a proper delegate selection. Explain. (If only the VP had taken an *Algorithms*, she could even have found a delegate selection without much effort.)

**Problem 11.9.**

A *Latin square* is $n \times n$ array whose entries are the number $1, \ldots, n$. These entries satisfy two constraints: every row contains all $n$ integers in some order, and also every column contains all $n$ integers in some order. Latin squares come up frequently in the design of scientific experiments for reasons illustrated by a little story in a footnote[10]

For example, here is a $4 \times 4$ Latin square:

| | | | |
|---|---|---|---|
| 1 | 2 | 3 | 4 |
| 3 | 4 | 2 | 1 |
| 2 | 1 | 4 | 3 |
| 4 | 3 | 1 | 2 |

**(a)** Here are three rows of what could be part of a $5 \times 5$ Latin square:

[10] At Guinness brewery in the eary 1900's, W. S. Gosset (a chemist) and E. S. Beavan (a "maltster") were trying to improve the barley used to make the brew. The brewery used different varieties of barley according to price and availability, and their agricultural consultants suggested a different fertilizer mix and best planting month for each variety.

Somewhat sceptical about paying high prices for customized fertilizer, Gosset and Beavan planned a season long test of the influence of fertilizer and planting month on barley yields. For as many months as there were varieties of barley, they would plant one sample of each variety using a different one of the fertilizers. So every month, they would have all the barley varieties planted and all the fertilizers used, which would give them a way to judge the overall quality of that planting month. But they also wanted to judge the fertilizers, so they wanted each fertilizer to be used on each variety during the course of the season. Now they had a little mathematical problem, which we can abstract as follows.

Suppose there are $n$ barley varieties and an equal number of recommended fertilizers. Form an $n \times n$ array with a column for each fertilizer and a row for each planting month. We want to fill in the entries of this array with the integers 1,...,$n$ numbering the barley varieties, so that every row contains all $n$ integers in some order (so every month each variety is planted and each fertilizer is used), and also every column contains all $n$ integers (so each fertilizer is used on all the varieties over the course of the growing season).

| | | | | |
|---|---|---|---|---|
| 2 | 4 | 5 | 3 | 1 |
| 4 | 1 | 3 | 2 | 5 |
| 3 | 2 | 1 | 5 | 4 |
| | | | | |
| | | | | |

Fill in the last two rows to extend this "Latin rectangle" to a complete Latin square.

**(b)** Show that filling in the next row of an $n \times n$ Latin rectangle is equivalent to finding a matching in some $2n$-vertex bipartite graph.

**(c)** Prove that a matching must exist in this bipartite graph and, consequently, a Latin rectangle can always be extended to a Latin square.

### Exam Problems

**Problem 11.10.**
Overworked and over-caffeinated, the Teaching Assistant's (TA's) decide to oust the lecturer and teach their own recitations. They will run a recitation session at 4 different times in the same room. There are exactly 20 chairs to which a student can be assigned in each recitation. Each student has provided the TA's with a list of the recitation sessions her schedule allows and no student's schedule conflicts with all 4 sessions. The TA's must assign each student to a chair during recitation at a time she can attend, if such an assignment is possible.

Describe how to model this situation as a matching problem. Be sure to specify what the vertices/edges should be and briefly describe how a matching would determine seat assignments for each student in a recitation that does not conflict with his schedule. This is a *modeling problem* —you need not determine whether a match is always possible.

**Problem 11.11.**
Because of the incredible popularity of Math for Computer Science, Rajeev decides to give up on regular office hours. Instead, each student can join some study groups. Each group must choose a representative to talk to the staff, but there is a staff rule that a student can only represent one group. The problem is to find a representative from each group while obeying the staff rule.

**(a)** Explain how to model the delegate selection problem as a bipartite matching problem.

**(b)** The staff's records show that no student is a member of more than 4 groups, and all the groups must have at least 4 members. That's enough to guarantee there is a proper delegate selection. Explain.

### Homework Problems

**Problem 11.12.**
Take a regular deck of 52 cards. Each card has a suit and a value. The suit is one of four possibilities: heart, diamond, club, spade. The value is one of 13 possibilities, $A, 2, 3, \ldots, 10, J, Q, K$. There is exactly one card for each of the $4 \times 13$ possible combinations of suit and value.

Ask your friend to lay the cards out into a grid with 4 rows and 13 columns. They can fill the cards in any way they'd like. In this problem you will show that you can always pick out 13 cards, one from each column of the grid, so that you wind up with cards of all 13 possible values.

**(a)** Explain how to model this trick as a bipartite matching problem between the 13 column vertices and the 13 value vertices. Is the graph necessarily degree-constrained?

**(b)** Show that any $n$ columns must contain at least $n$ different values and prove that a matching must exist.

**Problem 11.13.**
Scholars through the ages have identified *twenty* fundamental human virtues: honesty, generosity, loyalty, prudence, completing the weekly course reading-response, etc. At the beginning of the term, every student in Math for Computer Science possessed exactly *eight* of these virtues. Furthermore, every student was unique; that is, no two students possessed exactly the same set of virtues. The Math for Computer Science course staff must select *one* additional virtue to impart to each student by the end of the term. Prove that there is a way to select an additional virtue for each student so that every student is unique at the end of the term as well.

Suggestion: Use Hall's theorem. Try various interpretations for the vertices on the left and right sides of your bipartite graph.

## Problems for Section 11.6

### Practice Problems

**Problem 11.14.**
Four Students want separate assignments to four VI-A Companies. Here are their

preference rankings:

| Student | Companies |
|---|---|
| Albert: | HP, Bellcore, AT&T, Draper |
| Nick: | AT&T, Bellcore, Draper, HP |
| Oshani: | HP, Draper, AT&T, Bellcore |
| Ali: | Draper, AT&T, Bellcore, HP |

| Company | Students |
|---|---|
| AT&T: | Ali, Albert, Oshani, Nick |
| Bellcore: | Oshani, Nick, Albert, Ali |
| HP: | Ali, Oshani, Albert, Nick |
| Draper: | Nick, Ali, Oshani, Albert |

**(a)** Use the Mating Ritual to find *two* stable assignments of Students to Companies.

**(b)** Describe a simple procedure to determine whether any given stable marriage problem has a unique solution, that is, only one possible stable matching.

**Problem 11.15.**

We are interested in invariants of the Mating Ritual (Section 11.6) for finding stable marriages. Let Angelina and Jen be two of the girls, and Keith and Tom be two of the boys.

Which of the following predicates are invariants of the Mating Ritual no matter what the preferences are among the boys and girls? (Remember that a predicate that is always false is an invariant—check the definition of invariant to see why.)

**(a)** Angelina is crossed off Tom's list and she has a suitor that she prefers to Tom.

**(b)** Tom is serenading Jen.

**(c)** Tom is not serenading Jen.

**(d)** Tom's list of girls to serenade is empty.

**(e)** All the boys have the same number of girls left uncrossed in their lists.

**(f)** Jen is crossed off Keith's list.

**(g)** Jen is crossed off Keith's list and Keith prefers Jen to anyone he is serenading.

**(h)** Jen is the only girl on Keith's list.

## Class Problems

**Problem 11.16.**
Consider a stable marriage problem with 4 boys and 4 girls and the following partial information about their preferences:

| | | | | |
|---|---|---|---|---|
| B1: | G1 | G2 | – | – |
| B2: | G2 | G1 | – | – |
| B3: | – | – | G4 | G3 |
| B4: | – | – | G3 | G4 |
| G1: | B2 | B1 | – | – |
| G2: | B1 | B2 | – | – |
| G3: | – | – | B3 | B4 |
| G4: | – | – | B4 | B3 |

**(a)** Verify that

$$(B1, G1), (B2, G2), (B3, G3), (B4, G4)$$

will be a stable matching whatever the unspecified preferences may be.

**(b)** Explain why the stable matching above is neither boy-optimal nor boy-pessimal and so will not be an outcome of the Mating Ritual.

**(c)** Describe how to define a set of marriage preferences among $n$ boys and $n$ girls which have at least $2^{n/2}$ stable assignments.

*Hint:* Arrange the boys into a list of $n/2$ pairs, and likewise arrange the girls into a list of $n/2$ pairs of girls. Choose preferences so that the $k$th pair of boys ranks the $k$th pair of girls just below the previous pairs of girls, and likewise for the $k$th pair of girls. Within the $k$th pairs, make sure each boy's first choice girl in the pair prefers the other boy in the pair.

**Problem 11.17.**
Suppose there are more boys than girls.

**(a)** Define what a stable matching should mean in this case.

**(b)** Explain why applying the Mating Ritual in this case will yield a stable matching in which every girl is married.

### Homework Problems

**Problem 11.18.**
The most famous application of stable matching was in assigning graduating medical students to hospital residencies. Each hospital has a preference ranking of students and each student has a preference order of hospitals, but unlike the setup in the notes where there are an equal number of boys and girls and monogamous marriages, hospitals generally have differing numbers of available residencies, and the total number of residencies may not equal the number of graduating students. Modify the definition of stable matching so it applies in this situation, and explain how to modify the Mating Ritual so it yields stable assignments of students to residencies.

Briefly indicate what, if any, modifications of the preserved invariant used to verify the original Mating are needed to verify this one for hospitals and students.

**Problem 11.19.**
Give an example of a stable matching between 3 boys and 3 girls where no person gets their first choice. Briefly explain why your matching is stable.

**Problem 11.20.**
In a stable matching between $n$ boys and girls produced by the Mating Ritual, call a person *lucky* if they are matched up with one of their $\lceil n/2 \rceil$ top choices. We will prove:

**Theorem.** *There must be at least one lucky person.*

To prove this, define the following derived variables for the Mating Ritual:

$q(B) = j$, where j is the rank of the girl that boy $B$ is courting. That is to say, boy $B$ is always courting the $j$th girl on his list.

$r(G)$ is the number of boys that girl $G$ has rejected.

**(a)** Let

$$S ::= \sum_{B \in \text{Boys}} q(B) - \sum_{G \in \text{Girls}} r(G). \tag{11.4}$$

Show that $S$ remains the same from one day to the next in the Mating Ritual.

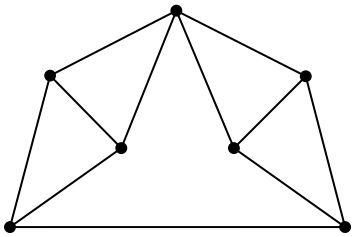
**(b)** Prove the Theorem above. (You may assume for simplicity that $n$ is even.)

*Hint:* A girl is sure to be lucky if she has rejected half the boys.

## Problems for Section 11.7

### Class Problems

**Problem 11.21.**
Let $G$ be the graph below[11]. Carefully explain why $\chi(G) = 4$.



### Homework Problems

**Problem 11.22.**
6.042 is often taught using recitations. Suppose it happened that 8 recitations were needed, with two or three staff members running each recitation. The assignment of staff to recitation sections, using their secret codenames, is as follows:

- R1: Maverick, Goose, Iceman
- R2: Maverick, Stinger, Viper
- R3: Goose, Merlin
- R4: Slider, Stinger, Cougar
- R5: Slider, Jester, Viper
- R6: Jester, Merlin
- R7: Jester, Stinger
- R8: Goose, Merlin, Viper

Two recitations can not be held in the same 90-minute time slot if some staff member is assigned to both recitations. The problem is to determine the minimum number of time slots required to complete all the recitations.

[11] From *Discrete Mathematics*, Lovász, Pelikan, and Vesztergombi. Springer, 2003. Exercise 13.3.1

**(a)** Recast this problem as a question about coloring the vertices of a particular graph. Draw the graph and explain what the vertices, edges, and colors represent.

**(b)** Show a coloring of this graph using the fewest possible colors. What schedule of recitations does this imply?

**Problem 11.23.**

This problem generalizes the result proved Theorem 11.7.3 that any graph with maximum degree at most $w$ is $(w+1)$-colorable.

A simple graph, $G$, is said to have *width*, $w$, iff its vertices can be arranged in a sequence such that each vertex is adjacent to at most $w$ vertices that precede it in the sequence. If the degree of every vertex is at most $w$, then the graph obviously has width at most $w$—just list the vertices in any order.

**(a)** Describe an example of a graph with 100 vertices, width 3, but *average* degree more than 5. *Hint:* Don't get stuck on this; if you don't see it after five minutes, ask for a hint.

**(b)** Prove that every graph with width at most $w$ is $(w+1)$-colorable.

**(c)** Prove that the average degree of a graph of width $w$ is at most $2w$.

**Problem 11.24.**

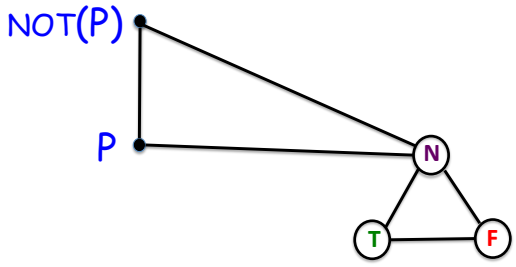
This problem will show that 3-coloring a graph is just as difficult as finding a satisfying truth assignment for a propositional formula. The graphs considered will all be taken to have three designated *color-vertices* connected in a triangle to force them to have different colors in any coloring of the graph. The colors assigned to the color-vertices will be called $T$, $F$ and $N$.

Suppose $f$ is an $n$-argument truth function. That is,

$$f : \{T, F\}^n \to \{T, F\}.$$

A graph $G$ is called a *3-color-f-gate* iff $G$ has $n$ designated *input vertices* and a designated *output vertex*, such that

- $G$ can be 3-colored *only* if its input vertices are colored with $T$'s and $F$'s.
- For every sequence $b_1, b_2, \ldots, b_n \in \{T, F\}$, there is a 3-coloring of $G$ in which the input vertices $v_1, v_2, \ldots, v_n \in V(G)$ have the colors $b_1, b_2, \ldots, b_n \in \{T, F\}$.




[h]

**Figure 11.26** A 3-coloring NOT-gate

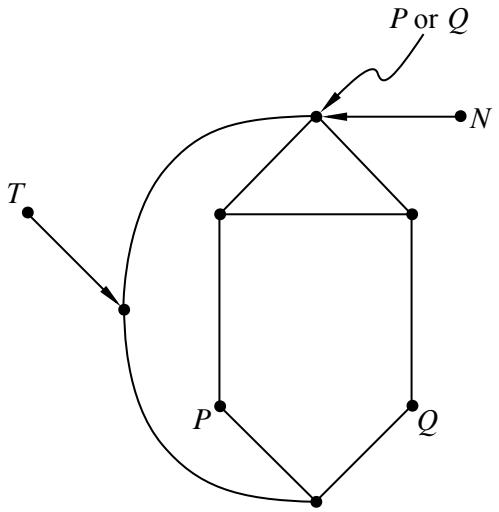
- In any 3-coloring of $G$ where the input vertices $v_1, v_2, \ldots, v_n \in V(G)$ have colors $b_1, b_2, \ldots, b_n \in \{T, F\}$, the output vertex has color $f(b_1, b_2, \ldots, b_n)$.

For example, a 3-color-NOT-gate consists simply of two adjacent vertices. One vertex is designated to be the input vertex, $P$, and the other is designated to be the output vertex. Both vertices have to be constrained so they can only be colored with $T$'s or $F$'s in any proper 3-coloring. This constraint can be imposed by making them adjacent to the color-vertex $N$, as shown in Figure 11.26.

**(a)** Verify that the graph in Figure 11.27 is a 3-color-OR-gate. (Color-vertex $F$ and the edges among the coloring vertices are not shown. Also not shown are edges from each of the input vertices $P$ and $Q$ to color-vertex $N$; these edges constrain $P$ and $Q$ to be colored $T$ or $F$ in any proper 3-coloring.)

**(b)** Let $E$ be an $n$-variable propositional formula, and suppose $E$ defines a truth function $f : \{T, F\}^n \to \{T, F\}$. Explain a simple way to construct a graph that is a 3-color-$f$-gate.

**(c)** Explain why an efficient procedure for determining if a graph was 3-colorable would lead to an efficient procedure to solve the satisfiability problem, SAT.




[h]

**Figure 11.27** A 3-coloring OR-gate

**Exam Problems**

**Problem 11.25.**

**False Claim.** *Let $G$ be a graph whose vertex degrees are all $\leq k$. If $G$ has a vertex of degree strictly less than $k$, then $G$ is $k$-colorable.*

**(a)** Give a counterexample to the False Claim when $k = 2$.

**(b)** Underline the exact sentence or part of a sentence that is the first unjustified step in the following bogus proof of the False Claim.

*Bogus proof.* Proof by induction on the number $n$ of vertices:

**Induction hypothesis**:

$P(n)$::= "Let $G$ be an $n$-vertex graph whose vertex degrees are all $\leq k$. If $G$ also has a vertex of degree strictly less than $k$, then $G$ is $k$-colorable."

**Base case**: ($n = 1$) $G$ has one vertex, the degree of which is 0. Since $G$ is 1-colorable, $P(1)$ holds.

**Inductive step**:

We may assume $P(n)$. To prove $P(n+1)$, let $G_{n+1}$ be a graph with $n+1$ vertices whose vertex degrees are all $k$ or less. Also, suppose $G_{n+1}$ has a vertex, $v$, of degree strictly less than $k$. Now we only need to prove that $G_{n+1}$ is $k$-colorable.

To do this, first remove the vertex $v$ to produce a graph, $G_n$, with $n$ vertices. Let $u$ be a vertex that is adjacent to $v$ in $G_{n+1}$. Removing $v$ reduces the degree of $u$ by 1. So in $G_n$, vertex $u$ has degree strictly less than $k$. Since no edges were added, the vertex degrees of $G_n$ remain $\leq k$. So $G_n$ satisfies the conditions of the induction hypothesis, $P(n)$, and so we conclude that $G_n$ is $k$-colorable.

Now a $k$-coloring of $G_n$ gives a coloring of all the vertices of $G_{n+1}$, except for $v$. Since $v$ has degree less than $k$, there will be fewer than $k$ colors assigned to the nodes adjacent to $v$. So among the $k$ possible colors, there will be a color not used to color these adjacent nodes, and this color can be assigned to $v$ to form a $k$-coloring of $G_{n+1}$. ■

**(c)** With a slightly strengthened condition, the preceding proof of the False Claim could be revised into a sound proof of the following Claim:

**Claim.** *Let $G$ be a graph whose vertex degrees are all $\leq k$. If* ⟨statement inserted from below⟩ *has a vertex of degree strictly less than $k$, then $G$ is $k$-colorable.*

Circle each of the statements below that could be inserted to make the proof correct.

- $G$ is connected and
- $G$ has no vertex of degree zero and
- $G$ does not contain a complete graph on $k$ vertices and
- every connected component of $G$
- some connected component of $G$

## Problems for Section 11.9

### Class Problems

**Problem 11.26.**

The $n$-dimensional hypercube, $H_n$, is a graph whose vertices are the binary strings of length $n$. Two vertices are adjacent if and only if they differ in exactly 1 bit. For example, in $H_3$, vertices `111` and `011` are adjacent because they differ only in the first bit, while vertices `101` and `011` are not adjacent because they differ at both the first and second bits.

**(a)** Prove that it is impossible to find two spanning trees of $H_3$ that do not share some edge.

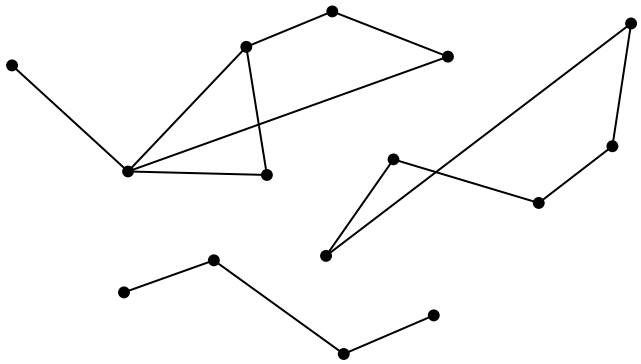
**(b)** Verify that for any two vertices $x \neq y$ of $H_3$, there are 3 paths from $x$ to $y$ in $H_3$, such that, besides $x$ and $y$, no two of those paths have a vertex in common.

**(c)** Conclude that the connectivity of $H_3$ is 3.

**(d)** Try extending your reasoning to $H_4$. (In fact, the connectivity of $H_n$ is $n$ for all $n \geq 1$. A proof appears in the problem solution.)

**Problem 11.27.**
A set, $M$, of vertices of a graph is a *maximal connected* set if every pair of vertices in the set are connected, and any set of vertices properly containing $M$ will contain two vertices that are not connected.

**(a)** What are the maximal connected subsets of the following (unconnected) graph?



**(b)** Explain the connection between maximal connected sets and connected components. Prove it.

**Problem 11.28. (a)** Prove that $K_n$ is $(n-1)$-edge connected for $n > 1$.

Let $M_n$ be a graph defined as follows: begin by taking $n$ graphs with non-overlapping sets of vertices, where each of the $n$ graphs is $(n-1)$-edge connected (they could be disjoint copies of $K_n$, for example). These will be subgraphs of $M_n$. Then pick $n$ vertices, one from each subgraph, and add enough edges between pairs of picked vertices that the subgraph of the $n$ picked vertices is also $(n-1)$-edge connected.

**(b)** Draw a picture of $M_4$.

**(c)** Explain why $M_n$ is $(n-1)$-edge connected.

**Problem 11.29.**

**False Claim.** *If every vertex in a graph has positive degree, then the graph is connected.*

**(a)** Prove that this Claim is indeed false by providing a counterexample.

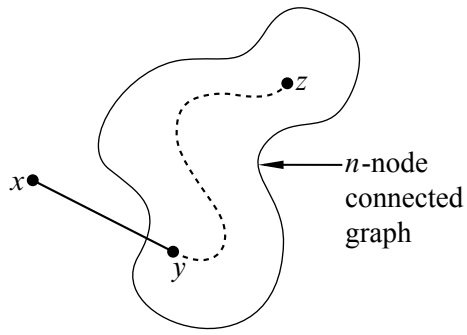
**(b)** Since the Claim is false, there must be an logical mistake in the following bogus proof. Pinpoint the *first* logical mistake (unjustified step) in the proof.

*Bogus proof.* We prove the Claim above by induction. Let $P(n)$ be the proposition that if every vertex in an $n$-vertex graph has positive degree, then the graph is connected.

**Base cases**: ($n \leq 2$). In a graph with 1 vertex, that vertex cannot have positive degree, so $P(1)$ holds vacuously.

$P(2)$ holds because there is only one graph with two vertices of positive degree, namely, the graph with an edge between the vertices, and this graph is connected.

**Inductive step**: We must show that $P(n)$ implies $P(n+1)$ for all $n \geq 2$. Consider an $n$-vertex graph in which every vertex has positive degree. By the assumption $P(n)$, this graph is connected; that is, there is a path between every pair of vertices. Now we add one more vertex $x$ to obtain an $(n+1)$-vertex graph:




All that remains is to check that there is a path from $x$ to every other vertex $z$. Since $x$ has positive degree, there is an edge from $x$ to some other vertex, $y$. Thus, we can obtain a path from $x$ to $z$ by going from $x$ to $y$ and then following the path from $y$ to $z$. This proves $P(n+1)$.

By the principle of induction, $P(n)$ is true for all $n \geq 0$, which proves the Claim.

■

### Homework Problems

**Problem 11.30.** **(a)** Give an example of a simple graph that has two vertices $u \neq v$ and two distinct paths between $u$ and $v$, but no cycle including either $u$ or $v$.

**(b)** Prove that if there are different paths between two vertices in a simple graph, then the graph has a cycle.

**Problem 11.31.**
The entire field of graph theory began when Euler asked whether the seven bridges of Königsberg could all be crossed exactly once. Abstractly, we can represent the parts of the city separated by rivers as vertices and the bridges as edges between the vertices. Then Euler's question asks whether there is a closed walk through the graph that includes every edge in a graph exactly once. In his honor, such a walk is called an *Euler tour*.

So how do you tell in general whether a graph has an Euler tour? At first glance this may seem like a daunting problem. The similar sounding problem of finding a cycle that touches every vertex exactly once is one of those Millenium Prize NP-complete problems known as the *Traveling Salesman Problem*). But it turns out to be easy to characterize which graphs have Euler tours.

**Theorem.** *A connected graph has an Euler tour if and only if every vertex has even degree.*

**(a)** Show that if a graph has an Euler tour, then the degree of each of its vertices is even.

In the remaining parts, we'll work out the converse: if the degree of every vertex of a connected finite graph is even, then it has an Euler tour. To do this, let's define an Euler *walk* to be a walk that includes each edge *at most* once.

**(b)** Suppose that an Euler walk in a connected graph does not include every edge. Explain why there must be an unincluded edge that is incident to a vertex on the walk.

In the remaining parts, let **w** be the *longest* Euler walk in some finite, connected graph.

**(c)** Show that if **w** is a closed walk, then it must be an Euler tour.

*Hint:* part (b)

**(d)** Explain why all the edges incident to the end of **w** must already be in **w**.

**(e)** Show that if the end of **w** was not equal to the start of **w**, then the degree of the end would be odd.

*Hint:* part (d)

**(f)** Conclude that if every vertex of a finite, connected graph has even degree, then it has an Euler tour.

### Homework Problems

**Problem 11.32.**
An edge is said to *leave* a set of vertices if one end of the edge is in the set and the other end is not.

**(a)** An $n$-node graph is said to be *mangled* if there is an edge leaving every set of $\lfloor n/2 \rfloor$ or fewer vertices. Prove the following:
**Claim.** *Every mangled graph is connected.*

An $n$-node graph is said to be *tangled* if there is an edge leaving every set of $\lceil n/3 \rceil$ or fewer vertices.

**(b)** Draw a tangled graph that is not connected.

**(c)** Find the error in the bogus proof of the following
**False Claim.** *Every tangled graph is connected.*

*Bogus proof.* The proof is by strong induction on the number of vertices in the graph. Let $P(n)$ be the proposition that if an $n$-node graph is tangled, then it is connected. In the base case, $P(1)$ is true because the graph consisting of a single node is trivially connected.

For the inductive case, assume $n \geq 1$ and $P(1), \ldots, P(n)$ hold. We must prove $P(n+1)$, namely, that if an $(n+1)$-node graph is tangled, then it is connected.

So let $G$ be a tangled, $(n+1)$-node graph. Choose $\lceil n/3 \rceil$ of the vertices and let $G_1$ be the tangled subgraph of $G$ with these vertices and $G_2$ be the tangled subgraph with the rest of the vertices. Note that since $n \geq 1$, the graph $G$ has a least two vertices, and so both $G_1$ and $G_2$ contain at least one vertex. Since $G_1$ and $G_2$ are tangled, we may assume by strong induction that both are connected. Also, since $G$ is tangled, there is an edge leaving the vertices of $G_1$ which necessarily connects to a vertex of $G_2$. This means there is a path between any two vertices of $G$: a path within one subgraph if both vertices are in the same subgraph, and a path traversing the connecting edge if the vertices are in separate subgraphs. Therefore, the entire graph, $G$, is connected. This completes the proof of the inductive case, and the Claim follows by strong induction.

■

**Problem 11.33.**
Let $G$ be the graph formed from $C_{2n}$, the cycle of length $2n$, by connecting every pair of vertices at maximum distance from each other in $C_{2n}$ by an edge in $G$.

**(a)** Given two vertices of $G$ find their distance in $G$.

**(b)** What is the *diameter* of $G$, that is, the largest distance between two vertices?

**(c)** Prove that the graph is not 4-connected.

**(d)** Prove that the graph is 3-connected.

## Problems for Section 11.11

### Practice Problems

**Problem 11.34. (a)** Prove that the average degree of a tree is less than 2.

**(b)** Suppose every vertex in a graph has degree at least $k$. Explain why the graph has a path of length $k$.

*Hint:* Consider a longest path.

### Exam Problems

**Problem 11.35.**
The $n$-dimensional hypercube, $H_n$, is a simple graph whose vertices are the binary strings of length $n$. Two vertices are adjacent if and only if they differ in exactly one bit. Consider for example $H_3$, shown in Figure 11.28. (Here, vertices `111` and `011` are adjacent because they differ only in the first bit, while vertices `101` and `011` are not adjacent because they differ in both the first and second bits.)

Explain why it is impossible to find two spanning trees of $H_3$ that have no edges in common.

### Class Problems

**Problem 11.36.**
Procedure *Mark* starts with a connected, simple graph with all edges unmarked and then marks some edges. At any point in the procedure a path that includes only marked edges is called a *fully marked* path, and an edge that has no fully marked path between its endpoints is called *eligible*.

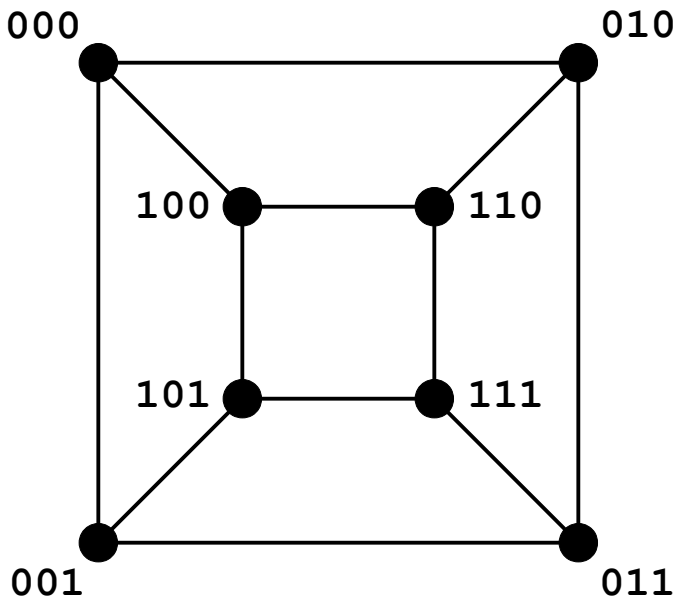



**Figure 11.28** $H_3$.

Procedure *Mark* simply keeps marking eligible edges, and terminates when there are none.

Prove that *Mark* terminates, and that when it does, the set of marked edges forms a spanning tree of the original graph.

**Problem 11.37.**

A procedure for connecting up a (possibly disconnected) simple graph and creating a spanning tree can be modelled as a state machine whose states are finite simple graphs. A state is *final* when no further transitions are possible. The transitions determined by the following rules:

**Procedure create-spanning-tree**

1. If there is an edge $\langle u—v\rangle$ on a cycle, then delete $\langle u—v\rangle$.
2. If vertices $u$ and $v$ are not connected, then add the edge $\langle u—v\rangle$.

**(a)** Draw all the possible final states reachable starting with the graph with vertices

$\{1, 2, 3, 4\}$ and edges

$$\{\langle 1\text{—}2\rangle, \langle 3\text{—}4\rangle\}.$$

**(b)** Prove that if the machine reaches a final state, then the final state will be a tree on the vertices graph on which it started.

**(c)** For any graph, $G'$, let $e$ be the number of edges in $G'$, $c$ be the number of connected components it has, and $s$ be the number of cycles. For each of the quantities below, indicate the *strongest* of the properties that it is guaranteed to satisfy, no matter what the starting graph is.

The choices for properties are: *constant*, *strictly increasing*, *strictly decreasing*, *weakly increasing*, *weakly decreasing*, *none of these*.

(i) $e$

(ii) $c$

(iii) $s$

(iv) $e - s$

(v) $c + e$

(vi) $3c + 2e$

(vii) $c + s$

**(d)** Prove that one of the quantities from part (c) strictly decreases at each transition. Conclude that for every starting state, the machine will reach a final state.

**Problem 11.38.**
Prove that a graph is a tree iff it has a unique path between every two vertices.

**Problem 11.39.**
Let $G$ be a weighted graph and suppose there is a unique edge $e \in E(G)$ with smallest weight, that is, $w(e) < w(f)$ for all edges $f \in E(G) - \{e\}$. Prove that any minimum weight spanning tree (MST) of $G$ must include $e$.

**Problem 11.40.**
Let $G$ be a $4 \times 4$ grid with vertical and horizontal edges between neighboring vertices. Formally,

$$V(G) = [0, 3]^2 ::= \{(k, j) \mid 0 \leq k, j \leq 3\}.$$

Letting $h_{i,j}$ be the horizontal edge $\langle (i, j)—(i+1, j)\rangle$ and $v_{j,i}$ be the vertical edge $\langle (j, i)—(j, i+1)\rangle$ for $i \in [0, 2]$, $j \in [0, 3]$. The weights of these edges are

$$w(h_{i,j}) ::= \frac{4i + j}{100},$$
$$w(v_{j,i}) ::= 1 + \frac{i + 4j}{100}.$$

(A picture of $G$ would help; you might like to draw one.)

**(a)** Construct a minimum weight spanning tree (MST) for $G$ by initially selecting the minimum weight edge, and then successively selecting the minimum weight edge that does not create a cycle with the previously selected edges. Stop when the selected edges form a spanning tree of $G$. (This is Kruskal's MST algorithm.)

**(b)** Grow an MST for $G$ starting with the tree consisting of the single vertex $(1, 2)$ and successively adding the minimum weight edge with exactly one endpoint in the tree. Stop when the tree spans $G$. (This is Prim's MST algorithm.)

**(c)** Grow an MST for $G$ by treating the vertices $(0, 0), (0, 3), (2, 3)$ as single vertex trees and then successively adding, for each tree in parallel, the minimum weight edge among the edges with one endpoint in the tree. Continue until the trees merge and form a spanning tree of $G$. (This is 6.042's parallel MST algorithm.)

**(d)** Verify that you got the same MST each time.

**Problem 11.41.**
In this problem you will prove:

**Theorem.** *A graph $G$ is 2-colorable iff it contains no odd length closed walk.*

As usual with "iff" assertions, the proof splits into two proofs: part (a) asks you to prove that the left side of the "iff" implies the right side. The other problem parts prove that the right side implies the left.

**(a)** Assume the left side and prove the right side. Three to five sentences should suffice.

**(b)** Now assume the right side. As a first step toward proving the left side, explain why we can focus on a single connected component $H$ within $G$.

**(c)** As a second step, explain how to 2-color any tree.

**(d)** Choose any 2-coloring of a spanning tree, $T$, of $H$. Prove that $H$ is 2-colorable by showing that any edge *not* in $T$ must also connect different-colored vertices.

### Homework Problems

**Problem 11.42.**
Prove Corollary 11.11.12: If all edges in a finite weighted graph have distinct weights, then the graph has a *unique* MST.

*Hint:* Suppose $M$ and $N$ were different MST's of the same graph. Let $e$ be the smallest edge in one and not the other, say $e \in M - N$, and observe that $N + e$ must have a cycle.